Virology 287, 112–132 (2001)

The Complete DNA Sequence of the *Ectocarpus siliculosus* Virus EsV-1 Genome

Nicolas Delaroque,*[,1] Dieter Gerhard Müller,* Gordana Bothe,† Thomas Pohl,† Rolf Knippers,* and Wilhelm Boland‡

*Universität Konstanz, Department of Biology, D-78457 Konstanz, Germany; †GATC Biotech AG, Jakob-Stadler-Platz, D-78467 Konstanz, Germany; and ‡Max-Planck-Institut für Chemische Oekologie, Carl Zeiss-Promenade 10, D-07745 Jena, Germany

Received May 5, 2001; returned to author for revision May 5, 2001; accepted May 18, 2001; published online July 23, 2001

The *Ectocarpus siliculosus* Virus-1, EsV-1, is the type-species of a genus of *Phycodnaviridae*, the phaeoviruses, infecting marine filamentous brown algae. The EsV-1 genome of 335,593 bp contains tandem and dispersed repetitive elements in addition to a large number of open reading frames of which 231 are currently counted as genes. Many genes can be assigned to functional groups involved in DNA synthesis, DNA integration, transposition, and polysaccharide metabolism. Furthermore, EsV-1 contains components of a surprisingly complex signal transduction system with six different hybrid histidine protein kinases and four putative serine/threonine protein kinases. Several other genes encode polypeptides with protein–protein interaction domains. However, 50% of the predicted genes have no counterparts in data banks. Only 28 of the 231 identified genes have significant sequence similarities to genes of the *Chlorella* virus PBCV-1, another phycodnavirus. To our knowledge, the EsV-1 genome is the largest viral DNA sequenced to date.

Key Words: Phycodnaviridae; *Ectocarpus siliculosus*; histidine kinase; phytochrome; integration; transposase; polysaccharide metabolism.

INTRODUCTION

Ectocarpus siliculosus Virus (EsV-1) is the founding member of a new genus of marine viruses, the phaeoviruses. They infect filamentous brown algae (phaeos, Greek: brown; Pringle, 1998). *E. siliculosus* is just one of a large number of marine brown algal species, and each of the eight host species that have been examined so far harbors its own species-specific phaeovirus (Müller *et al.,* 1998). Phaeoviruses have an icosahedral morphology with a dense nucleoprotein core surrounded by a multilayered protein coat. Their genetic material is large double-stranded DNA, a feature that phaeoviruses share with other viruses of the phycodnaviridae family whose most prominent members are the well-studied plaque-forming *Chlorella* viruses (Van Etten *et al.,* 1991).

However, in contrast to *Chlorella* viruses with their classic lytic infection cycle (Van Etten and Meints, 1999), EsV-1 and other phaeoviruses are lysogenic. They initiate their life cycle by infecting the free-swimming, wall-less gametes or spores of their algal hosts. The viral DNA becomes integrated into the host genome and is transmitted to all cells of the developing alga (Müller, 1991a; Bräutigam *et al.,* 1995; Delaroque *et al.,* 1999). The viral genome remains latent in vegetative cells, but is expressed in cells of the reproductive algal organs, sporangia, and gametangia. Massive replication of viral DNA occurs in hypertrophic nuclei and is followed by nuclear breakdown and viral assembly which continues until the cell is densely packed with viral particles (Lanka *et al.,* 1993; Müller *et al.,* 1998; Wolf *et al.,* 1998, 2000). Virions are released into the surrounding sea water under conditions that are also optimal for the release of spores or gametes such as changes in temperature, light, and sea water composition (Müller, 1991b). This synchronization facilitates an interaction of viruses with their susceptible host cells. EsV-1 is pandemic in *E. siliculosus* populations on the coasts of all oceans in the temperate climate zones (Sengco *et al.,* 1996; Müller *et al.,* 2000).

A detailed investigation of EsV-1 is expected to be rewarding for several reasons. One is that EsV-1 could serve as a probe to investigate the molecular biology of its host. For example, the cell-specific replication of viral DNA implies that cells in reproductive organs express factors essential for viral gene expression. It is likely that these factors have not evolved to end the period of viral latency, but to participate in the induction or maintenance of the differentiated state of gametangia and sporangia. The identification of these factors would therefore give important insights into the developmental program of reproduction in marine algae. It can also be expected that EsV-1 will be useful in biotechnology (Henry and Meints, 1994; Müller *et al.,* 1998). Integrated latent viral DNA carrying a foreign gene should produce high quantities of the product of that gene. This could be technically exploited because latently infected algae remain somatically unaffected and can be mass-cultivated at low cost.

[1] To whom correspondence and reprint requests should be addressed. Fax: +49 7531 88 2966. E-mail: Nicolas.Delaroque@uni-konstanz.de.

0042-6822/01 $35.00

There are also ecological reasons for further studies on EsV-1 since this virus is widely distributed within all *Ectocarpus* populations worldwide. Fifty to 100% of all specimens within an *Ectocarpus* population usually carry the latent viral genome (Sengco *et al.,* 1996; Müller *et al.,* 2000). To explain this high degree of infection, it has been proposed that EsV-1 evolved to reduce sexual recombination in favor of mitotic reproduction in its algal host (Müller *et al.,* 1998), but this possibility has yet to be substantiated by experiments.

Lanka *et al.* (1993) first showed that the viral genome is a large DNA molecule. Using pulsed-field agarose gel electrophoresis, they detected the extracted virus DNA in three fractions: a wide spectrum of DNA fragments in a size range of 10–60 kb; a prominent band of linear DNA of about 330 kb in length; and a second prominent DNA band that remained close to the origin of the gel. Electron microscopy and restriction fragment analyses established that a major portion of EsV-1-DNA occurs in the form of an extended circle which does not enter the agarose gel during pulsed-field electrophoresis. The authors concluded that the natural form of the viral genome is a double-stranded DNA circle that tends to break, giving rise to full-length linear DNA forms as well as DNA fragments. An additional feature of EsV-1-DNA is multiple single-stranded regions (Lanka *et al.,* 1993), some of which may occur at specific sites in the genome (Klein *et al.,* 1994).

Only a few EsV-1 genes have been isolated and described so far. The gene gp1 encodes one of the three known structural glycoproteins (Klein *et al.,* 1995), and genes vp55 and vp74 code for major core proteins (Delaroque *et al.,* 2000a). A surprising discovery was that another constituent of the viral particle is encoded by a gene, vhk-1, with high sequence similarities to bacterial and plant genes for histidine protein kinases, elements of two-component signal transduction pathways. This was the first time that a regulatory gene of this type was detected in a viral system (Delaroque *et al.,* 2000b).

We have continued our analysis of the EsV-1 genome and report now the entire sequence of 335,593 bp that constitute the viral DNA. The genome contains 231 major open reading frames and a number of interspersed regions with repetitive DNA elements. The EsV-1 genome is the second phycodnavirus genome to be sequenced; the first is the *Chlorella* virus PBCV-1 genome (Van Etten and Meints, 1999). As expected, both viruses share a set of genes, but differ in genome organization and in a large number of genes that are unique to one of the two viruses. Comparisons give interesting insights into the evolution of algal viruses.

RESULTS AND DISCUSSION

Description of the viral genome

The viral DNA consists of 335,593 bp with 48.3% A+T nucleotides in good agreement with previous estimates based on pulsed-field gel electrophoresis and chromatographic separations of nucleotides (Lanka *et al.,* 1993). The percentage of A+T nucleotides is lower than the A+T content of other large DNA viruses such as chloroviruses (Van Etten and Meints, 1999), asfarviruses (Yáñez *et al.,* 1995), baculoviruses (Kuzio *et al.,* 1999), and poxviruses (Bawden *et al.,* 2000).

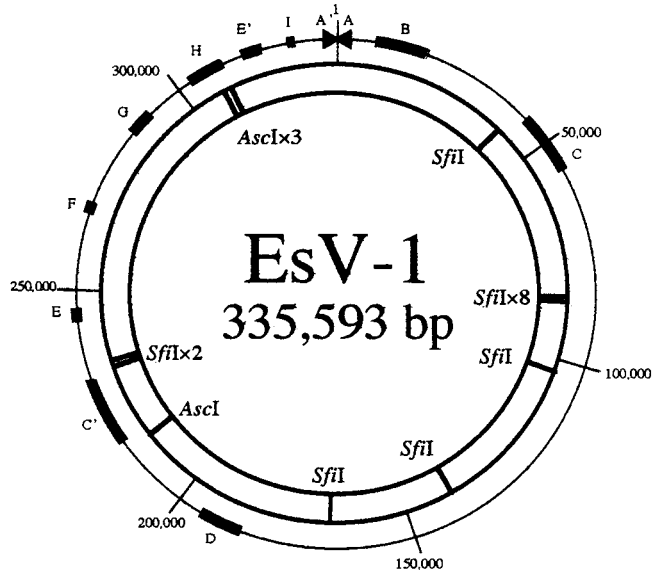

FIG. 1. Circular map of the EsV-1 genome. Inner circle, sites for restriction endonucleases *Asc*I and *Sfi*I. Outer circle, nucleotide coordinates and position of repeat regions (block rectangles: B, C, C′, etc.). Triangles, the inverted terminal repeats, ITRs A and A′.

Inverted repeats at the genome ends. Alignments of individually sequenced overlapping DNA fragments resulted in one large contig leaving one gap that is bordered on both sides by almost perfect inverted repeats of 1.800 and 1.560 bp in length. Several attempts failed to close this gap by PCR reactions using primers corresponding to sequences within the inverted repeat regions. We therefore conclude that the inverted repeats mark the ends of the viral genome and designate the first nucleotide in the "right" inverted repeat as the start of the sequence map and the last nucleotide in the "left" inverted repeat as its end (ITR, inverted terminal repeats A and A′ in Fig. 1).

A linear map appears to be at odds with the earlier results which had been interpreted to indicate that the EsV-1 genome is circular (see Introduction; Lanka *et al.,* 1993). To reconcile these apparently conflicting data we propose as one possibility that the complementary sequences in the right ITR-A and in the left ITR-A′ anneal to form a cruciform structure which effectively closes the DNA circle (Fig. 2A). In fact, a reexamination of earlier restriction maps shows that *Asc*I digestion gives two major bands, 244 and 90 kb long, compatible with a circular map, and two weaker bands of 220 and 26 kb that, together with the 90-kb fragment, can be arranged to give a linear DNA molecule ending in ITR-A and ITR-A′

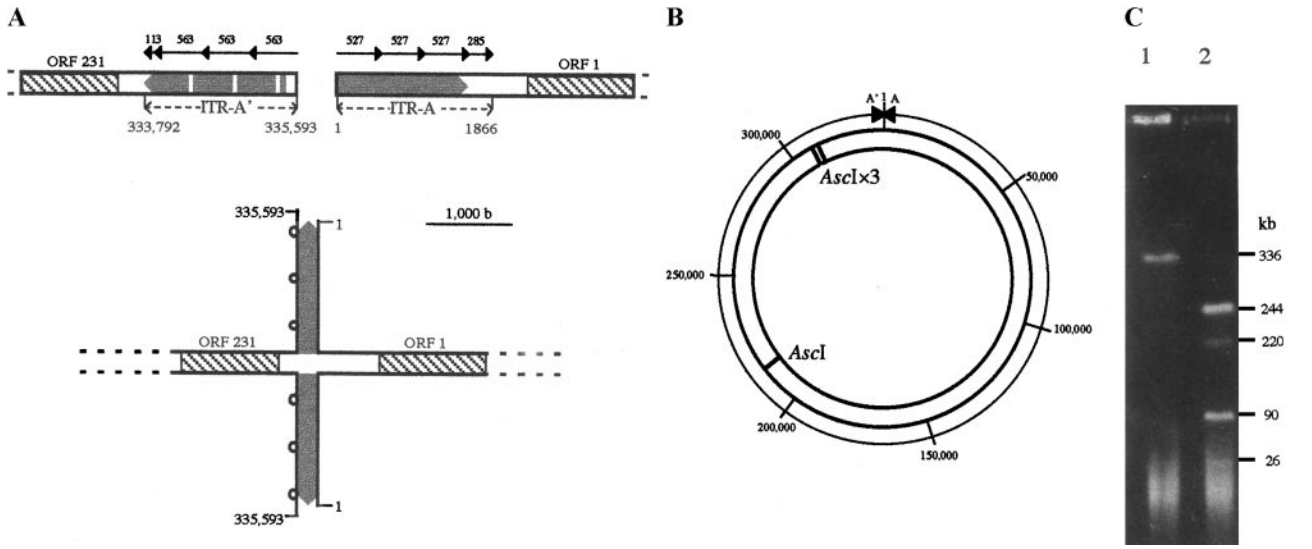

FIG. 2. A possible cruciform DNA conformation at the site of ring closure. (A) Scheme of the inverted terminal repeats. Upper, the identical 527-bp repeats are continuous in ITR-A, but interrupted by 36-bp gaps in ITR-A′. Lower, the inverted repeat sequences are aligned forming a cruciform DNA structure. (B) EsV-1 map with *Asc*I restriction sites. (C) Pulsed-field agarose gel electrophoresis of unrestricted EsV-1 DNA (lane 1) and *Asc*I-digested EsV-1 DNA (lane 2). The size of the DNA fragments are estimated according to concatemeric λ-DNA marker (not shown, but see Lanka *et al.,* 1993). Note that unrestricted EsV-1 DNA partitions to three positions in the gel: at the start (circular DNA), at 335 kb (linear DNA), and in the form of DNA fragments.

(Figs. 2B and 2C). Furthermore, restriction of the EsV-1 DNA recovered from the slot of the pulsed-field gel (Fig. 2C, lane 1) gave fragments compatible with a circular map, whereas restriction of the freely migrating EsV-1 DNA of unit length resulted in restriction fragments that can be composed as a linear map (not shown). This biochemical evidence would be consistent with the notion that the ends of the EsV-1 genome may be linked by cross-annealing of end sequences. However, it is also possible that the ends of a fraction of the EsV-1 genomes as isolated from viral particles are covalently closed analogous to lambda and other temperate phages, while another fraction of the EsV-1 genomes remains linear.

The right inverted repeat, ITR-A, consists of three copies of a 527-bp repeat and a fourth copy of 285 bp corresponding to a segment of the unit length repeats (Fig. 2A). Likewise, the left inverted repeat, ITR-A′, also contains three identical 527-bp copies followed by a 113-bp incomplete copy. The repeat elements in ITR-A′ are interrupted by 36-bp gap sequences. Therefore, if terminal cruciforms occur, they are predicted to contain short single-stranded loops in addition to 130-bp-long protruding single strands of sequences present in ITR-A′, but not in ITR-A (Fig. 2A).

Repetitive elements. The EsV-1 genome contains a number of regions with tandem repeats. An example is region B (see Fig. 1), which is mainly composed of 11 tandemly oriented segments of 330 and 260 bp lengths. Similarly, region C and the related region C′ (Fig. 1) contain numerous 65-bp elements, separated by a 562-bp single-copy element in region C (Fig. 3). In addition, region C′ possesses thirteen 120-bp repeats. All repeats in regions C and C′ begin with imperfect palindromes of 49 bp (Fig. 3). Upstream of these repeats in region C′ are four unrelated 186-bp tandem repeats embedded in a sequence context of unusually high G+C content (61%, not shown). Arrays of direct repeats are also found in region D, G, and I of the EsV-1 genome (Fig. 1; see also below, Fig. 5).

The function of these regions with tandem repeats remains to be elucidated. We note though that other large viral genomes such as that of baculoviruses also contain multicopy tandem repeat regions interspersed along the genome (Kuzio *et al.,* 1999). The baculovirus repeats are known to be important elements of transcriptional enhancers and origins of viral DNA replication (Kuzio *et al.,* 1999 and references therein). A possibility is that the tandem repeats in EsV-1 also serve as origins and remain in an unwound, single-stranded conformation upon packaging into the viral coat. This could explain the reason for the extensive single-strandedness in extracted viral DNA (Klein *et al.,* 1994).

Additional EsV-1 repeat regions consist of longer DNA elements. Regions E and E′ (Fig. 1) are similar and are composed each of two inverted ca. 2.600-bp sequences which could encode a functional protein (see below). Similarly, region F includes two inverted ca. 800-bp elements with coding potential.

Together, the repeat regions occupy approximately 12% of the EsV-1 genome, and an additional 22% of the genome have no apparent coding potential such as region H (Figs. 1 and 5). Thus, only two thirds of the EsV-1 genome may contain functional genes.

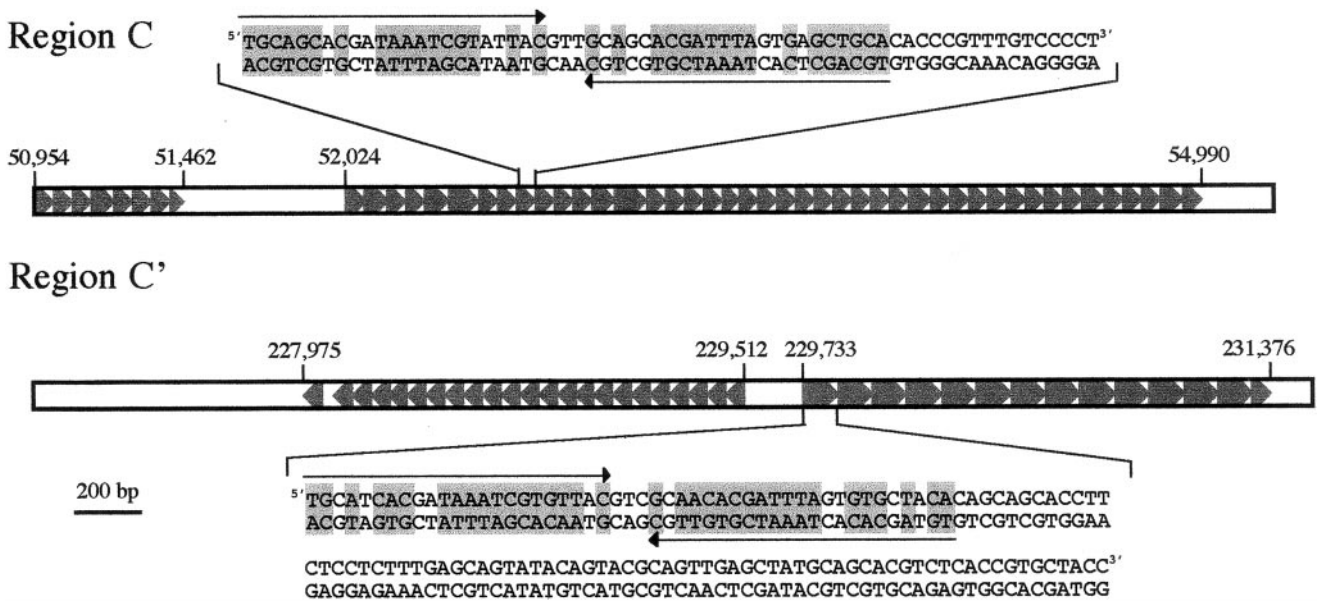

FIG. 3. Organization of the repeat regions C (top) and C′ (bottom). Repeats are represented as gray arrows. Region C mainly contains 65-bp direct repeats separated by a 562-bp gap, whereas region C′ is composed of one array with 65-bp repeats and a second array with 120-bp repeats. Two repeats with their imperfect palindromic elements (arrows) are shown.

Genes

We define open reading frames (ORF) as potential genes when they fulfill two criteria. First, they must be equal or longer than 65 codons, and second they must possess an upstream AT-rich element. The second criterion is based on our observation that all four analyzed EsV-1 genes have AT-rich sequences within 50 bp upstream of their putative ATG initiation codons (Fig. 4). This is also a characteristic feature of the major ORFs in the *Chlorella* virus PBCV-1 genome (Schuster *et al.,* 1990), suggesting that these sites have an important function for the expression of phycodnavirus genes. As an exception to these criteria, we count ORFs as genes when they have the potential to encode a protein with similarities to proteins in data banks. Following these criteria the EsV-1 genome contains 231 major genes, including those for structural proteins that have been previously identified (Klein *et al.,* 1995; Delaroque *et al.,* 2000a,b).

It is quite possible that our criteria are too strict. For example, the relatively large ORFs in three dispersed repeat elements (R1–R3; see below) are not included in the list of EsV-1 genes because they do not contain an upstream stretch of AT-base pairs and show no similarities to known gene sequences. Moreover, we expect that the rapidly expanding data banks will soon contain entries that could match several of the unassigned ORFs in the EsV-1 genome and that future investigations of the virus' molecular biology may uncover functional genes that we did not recognize in this first analysis of the viral coding potential.

Presently, we have no indication for introns in potential protein-coding genes, and we have no evidence for tRNA-, rRNA-, or other RNA-coding genes. One hundred eleven ORFs (48%) could encode proteins with similarities to sequences in data bases. Of these only 28 are similar to genes of the other sequenced phycodnavirus DNA, the *Chlorella* virus PBCV-1 genome.

An overview of the viral gene loci is given in Fig. 5. The selected 231 major genes are numbered according to their position in the virus genome irrespective of their orientation. In fact, 133 genes are transcribed in a clock-

```
gp1       GCGAGCTTT AAATAGTATGTACTTATGAA---------------------------- ATG
                                                                    first M
vp55           CAA AAATAAAATGTTGCGT GAGCTTAATTGATATATATACTGAAAACCAAA ATG
vp27           GCG AAATATTATGTAGACT TGAATAAAAG---------------------- ATG
vp74           ATA AAATATAATGTTGACA ACACCCAAACAAAGATAA-------------- ATG
vhk-1    AACGTGTGT GAAAAAAATATTGACA TGTGTACAGGAATGCGTGC------------- ATG
gp1   (48)-ATCGTACA AAATAAAATATTGCTT GAATATAAAAAC-------------------- ATG
                                                                   second M
```

FIG. 4. A/T-rich upstream elements. Alignment of regions upstream of the ATG initiation codons in genes gp1 (Klein *et al.,* 1995), vp55, vp27, vp74 (Delaroque *et al.,* 2000a), and vhk-1 (Delaroque *et al.,* 2000b). The A/T-rich elements are indicated by the gray boxes.

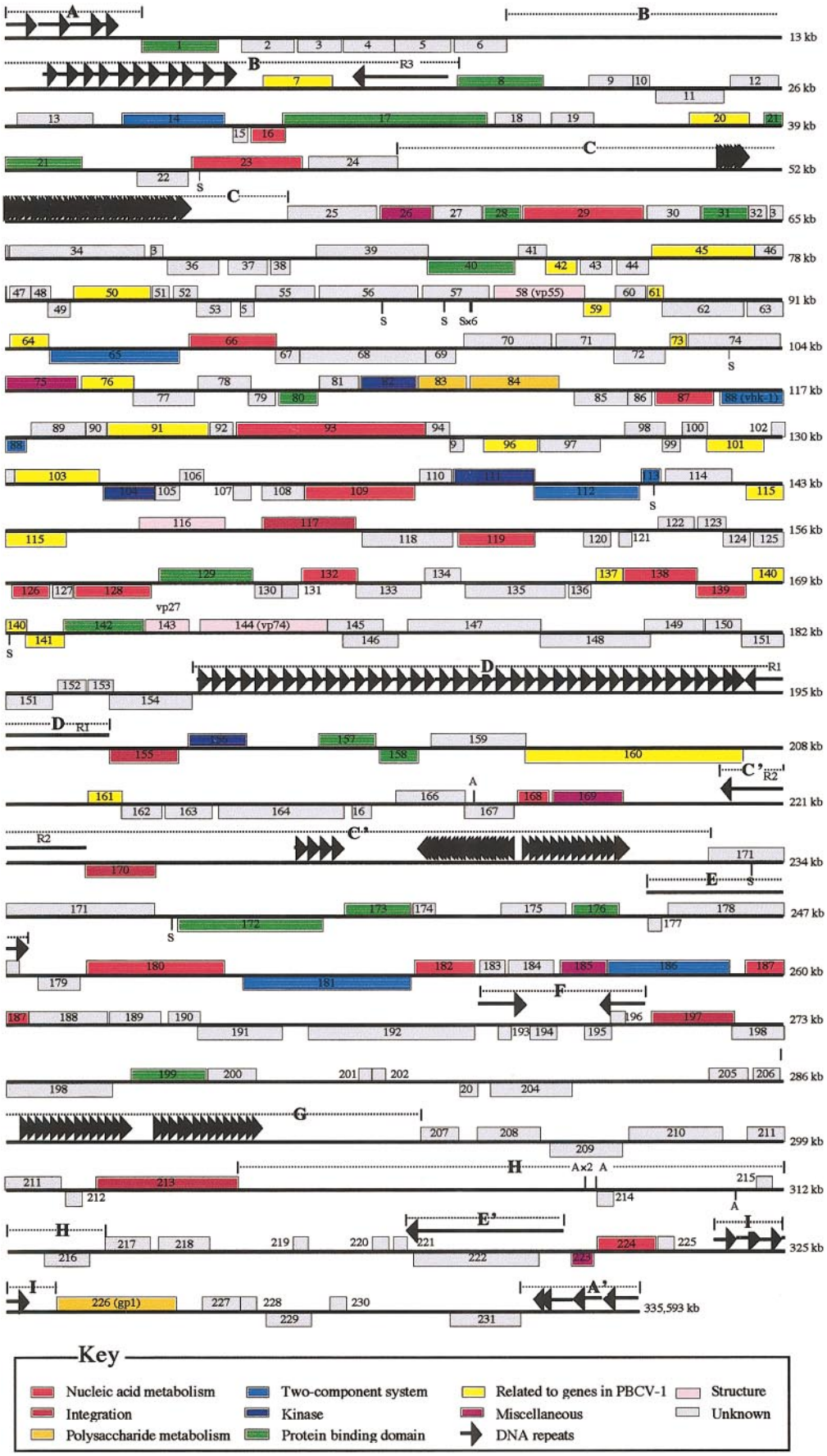

FIG. 5. Gene map of the EsV-1 genome. ORFs transcribed rightward are located above the line and ORFs transcribed leftward are below the line. Genes with similar functions are colored according to the key. Large repeats are represented as arrows. The *Asc*I and *Sfi*I restriction sites are indicated as A and S, respectively. For details, see Table 1.

TABLE 1

The Coding Potential of the EsV-1 Genome

ORF	Start	Stop	Size (aa)	Blast (E value)	amino acid identity	Homolog(s) [accession number][a]	Putative function/Feature(s)[b]
1	3552	2284	422	6.4e-19 1.2e-18	74/234 (31%) 98/315 (31%)	*Macaca fascicularis* ankyrin product (length = 307) [AB046089] Human KIAA1223 ankyrin product (length = 768) [AB033049]	7 ankyrin repeats (aa 50-82, 83-115, 125-157, 158-190, 194-223, 251-284, 292-324) [PF00023] [PS50297] [ANK], Serine protease, subtilase family, aspartic acid active site (aa 241-253) [PS00136]
2	4804	3947	285				
3	5590	4883	235				
4	6487	5651	278				
5	7426	6509	305				CC (aa 27-56)
6	8371	7514	285				
7	17330	18466	378	1.1e-15 1.8e-15 5.7e-09	66/202 (32%) 62/185 (33%) 56/195 (28%)	Chlorovirus PBCV-1 hypothetical protein A564L (length = 351) [U42580] Chlorovirus PBCV-1 hypothetical protein A154L (length = 347) [U42580] Chlorovirus PBCV-1 hypothetical protein A129R (length = 358) [U42580]	Leu-rich region (aa 319-377) [PS50319], NCC (aa 238-266)
8	20592	22004	470	1.7e-26 2.9e-24	122/399 (30%) 110/356 (30%)	Human antigen NY-BR-16 (length = 1188) [AF308285] Human Ankyrin protein (length = 435) [AK022041]	7 ankyrin repeats (aa 40-72, 74-106, 116-148, 190-218, 219-251, 252-284, 344-378) [PF00023] [PS50297] [ANK]
9	22803	23495	230				
10	23503	23778	91				Pro-rich region (aa 75-82) [PS50099], TM (aa 19-39)
11	25025	23910	371				Arg-rich region (aa 3-48) [PS50323], MCC (aa 3-48)
12	25141	25944	267	0.00024	25/65 (38%)	*Bacillus halodurans* BH2857 protein (length = 405) [AP001516]	
13	26218	27459	413				
14	27958	29679	573	9e-15 1e-16 2.0e-16	108/415 (26%) 105/424 (24%) 110/413 (26%)	*Xanthomonas campestris* RPFC histidine kinase (length = 677) [S16003] *Xanthomonas oryzae* RPFC histidine kinase (length = 676) [X97865] *Xylella fastidiosa* regulator of pathogenicity factors (length = 662) [AE003947]	**Hybrid histidine kinase** Histidine kinase domain (aa 199-421) [PF00512] [PS50109] [HisKA] [HATPase_c], Response regulator receiver domain (aa 445-562) [PF00072] [PS50110] [REC], SP (aa 1-24)
15	29951	29727	74				SP (aa 1-26), TM (aa 39-59)
16	30584	30018	188	0.0028 0.23	16/45 (35%) 18/59 (30%)	Bacteriophage phi-C31 dnase (length = 115) [X91149] *Escherichia coli* 5-methylcytosine-specific restriction enzyme A (length = 277) [P24200]	**Endonuclease** HNH endonuclease (aa 56-99) [PF01844] [HNHc], Prokaryotic membrane lipoprotein lipid attachment site (aa 176-186) [PS00013]
17	30668	34069	1133	6.0e-20 4.6e-19	72/215 (33%) 77/194 (39%)	*Chromatium vinosum* ankyrin homolog precursor (length = 323) [L13419] Human KIAA1223 protein (length = 768) [AB033049]	5 ankyrin repeats (aa 49-81, 82-114, 115-151, 152-185, 186-220) [PF00023] [PS50297] [ANK]
18	34197	34925	242	0.67 0.98	32/112 (28%) 32/112 (28%)	*Serratia entomophila* RNA polymerase sigma factor (length = 332) [U35777] *Salmonella typhimurium* LT2 mutant RPOS (length = 330) [AF184103]	CC (aa 32-59)
19	35146	35844	232	0.0048 0.028 0.062	22/44 (50%) 25/87 (28%) 23/79 (29%)	Bovine herpesvirus type 1trans-acting transcriptional protein ICP0 (p135 protein) (length = 676) [M84465] Rabbit fibroma virus S143R product (length = 234) [L26342] Myxoma virus M143R product (length = 234) [AF170726]	Ring finger (aa 109-149) [PS50089] [RING], Cytochrome c family heme-binding site signature (aa 125-130) [PS00190], Growth factor and cytokines receptors family signature 1 (aa 125-138) [PS00241], Cys-rich region (aa 125-148) [PS50311]
20	37470	38471	333	1.0e-107 0.0016 0.037 0.038	178/330 (53%) 30/121 (24%) 22/74 (29%) 22/58 (37%)	EsV-1-208 *Arabidopsis thaliana* (length = 208) [AB017069] *Mus musculus* TRIF-D (length = 95) [AB025011] Chlorovirus PBCV-1 hypothetical protein A481L (length = 224) [U42580]	Ring finger (aa 126-173) [PF00097] [PS50089] [RING]
21	38711	40282	523	2.3e-19 1.1e-18 1.3e-18	107/342 (31%) 106/342 (30%) 107/342 (31%)	Human ankyrin G119 (length = 1088) [U43965] *Mus musculus* ankyrin G (length = 1943) [L40631] Human ankyrin G (length = 4377) [U13616]	9 ankyrin repeats (aa 80-110, 115-147, 150-182, 187-219, 252-284, 295-327, 328-360, 361-390, 395-429) [PF00023] [PS50297] [ANK], Prokaryotic membrane lipoprotein lipid attachment site (aa 45-55) [PS00013], MCC (aa 443-497), TMs (aa 40-60, 329-349, 496-516)
22	42073	41219	284				ATP/GTP-binding site motif A (aa 233-240) [PS00017], Ser-rich regions (aa 49-140, 234-255) [PS00017], MCC (aa 7-35)
23	42136	43974	612	1.8e-12	109/474 (22%)	*Bacillus subtilis* hypothetical helicase in SINI-GCVT intergenic region (length =	**Helicase** Helicases conserved C-terminal domain

Abbreviations: TM, transmembrane domain; SP, signal peptide; MCC, mixed charge cluster; PCC, positive charge cluster; NCC, negative charge cluster; CC, coiled-coil domain.

[a] Accession numbers are from the GenBank database.

[b] Pfam, prosite or Smart codes are in brackets (E-value < 1) as well as the accession number for the known EsV-1 proteins.

TABLE 1—*Continued*

				7.5e-06 2.4e-05	105/411 (25%) 73/319 (22%)	557) [D84432] *Deinococcus radiodurans* helicase, SNF2/RAD54 family (length = 600) [AE001973] *Bacillus halodurans* SNF2 helicase (length = 995) [AP001517]	(aa 464-545) [PF00271] [PS50136] [HELICc], DEAD/DEAH box helicase (aa 244-324) [PF00270] [DEXDc], ATPases associated with various cellular activities (aa 164-526) [PF00004] [AAA], ATP/GTP-binding site motif A by similarity (aa 169-176), CC (aa 29-63)
24	44103	45572	489				Regulator of chromosome condensation (RCC1) signature 2 (aa 25-35) [PS00626], Pro-rich region (aa 429-472) [PS50099], Leu-rich regions (aa 123-143, 431-486) [PS50319], TMs (aa 123-143, 355-375, 402-422, 465-485)
25	56751	58229	492				Fibronectin type III domain (aa 2-71) [PF00041] [FN3], Eukaryotic putative RNA-binding region RNP-1 signature (aa 317-324) [PS00030], Proteins binding FMN and related compounds (aa 210-245) [PS50264]
26	58289	59161	290	1.1e-34 2.1e-26	81/233 (34%) 77/233 (33%)	Chlorovirus PBCV-1 ATPase (A392R protein, length = 258) [U42580] Chilo iridescent virus ATPase (length = 258) [AF003534]	**ATPase** ATPases associated with various cellular activities (aa 30-183) [AAA], ATP/GTP-binding site motif A (aa 38-45) [PS00017]
27	59187	59984	265				TMs (aa 27-47, 54-74, 87-107, 115-135, 143-163)
28	60018	60632	204				SET-domain of transcriptional regulators (aa 3-170) [PF00856] [PS50280] [SET]
29	60675	62702	675	4.0e-08 2e-07 1.1e-05 2.1e-05	67/286 (23%) 57/234 (24%) 77/316 (24%) 48/189 (25%)	*Mus musculus* DNA helicase B (length = 1175) [AB048542] *Methanococcus jannaschii* hypothetical protein MJ1519 (length = 1175) [U67593] *Chlamydia pneumoniae* RecD (length = 732) [AE001598] *Mycobacterium smegmatis* RecD (length = 554) [AF157643]	**Helicase** Viral (Superfamily 1) RNA helicase (aa 610-648) [PF01443], ATPases associated with various cellular activities (aa 268-482) [AAA], ATP/GTP-binding site motif A (aa 276-283) [PS00017]
30	62756	63631	291				
31	63677	64435	252	2.6e-18 3.7e-16	60/189 (31%) 67/212 (31%)	Fowlpox virus ankyrin FPV245 (length = 436) [AF198100] Human ankyrin G119 (length = 1088) [U43965]	4 ankyrin repeats (aa 31-60, 61-90, 91-120, 121-150) [PF00023] [PS50297] [ANK]
32	64472	64738	88				SP (aa 1-27)
33	64811	65068	85	0.0017	22/56 (39%)	*Helicobacter pylori* hypothetical 18.0 kDa protein (length = 149) [AE000638]	Lys-rich region (aa 39-84) [PS50318], MCC (aa 38-74), CC (aa 35-74), SP (aa 1-24)
34	65100	67328	742	2.2e-05 2.3e-05 9.2e-05	31/109 (28%) 82/350 (23%) 53/260 (20%)	*Rattus norvegicus* periaxin (length = 1389) [Z29649] *Bacillus halodurans* phage-related protein (length = 940) [AP001519] *Methanobacterium* phage psiM2 tail protein (length = 1186) [AF065411]	TMs (aa 20-40, 534-554, 643-663, 676-696, 699-719)
35	67449	67649	66				SP (aa 1-39)
36	68606	67758	282				MCC (aa 247-278), CC (aa 253-279)
37	69413	68781	210	0.00047	38/131 (29%)	*Arabidopsis thaliana* T09D09.15 product (length = 269) [AC002338]	Pro-rich region (aa 91-109) [PS50099], Ser-rich region (aa 108-155) [PS50324]
38	69792	69478	104				
39	70239	72080	613				Fibronectin type III domain (aa 388-467) [PF00041], TM (aa 31-51, 120-140, 400-420, 552-572)
40	73547	72084	487	0.30	35/134 (26%)	Human LZTR-1 (length = 552) [D38496]	BTB/POZ domain (aa 18-143) [PF00651] [PS50097]
41	73599	74072	157				
42	74593	74069	174	3.4e-17 8.3e-07	57/174 (32%) 30/106 (28%)	Chlorovirus PBCV-1 hypothetical protein A408L (length = 277) [U42580] Chlorovirus PBCV-1 hypothetical protein A410L (length = 110) [U42580]	Importin-beta N-terminal domain (aa 86-111) [PS50166], TM (aa 104-124)
43	75183	74653	176				SP (aa 1-35), TM (aa 15-35)
44	75773	75258	171	1.3e-09	36/106 (33%)	*Saccharomyces cerevisiae* suppressor protein SRP40 (length = 406) [L11275]	Asp-rich region (aa 65-170) [PS50312], Ser-rich region (aa 102-128) [PS50324], PCC (aa 41-61), TM (aa 17-37)
45	75843	77543	566	5.0e-09	73/274 (26%)	Chlorovirus PBCV-1 hypothetical protein A468R (length = 443) [U42580]	
46	77588	78049	153				Ala-rich region (aa 65-100) [PS50310], TM (aa 32-52)
47	78090	78410	106				SP (aa 1-25)
48	78472	78753	93				

wise and 98 in a counterclockwise direction relative to the map as defined in Fig. 1.

The major EsV-1 genes with their coding potential and the results of database searches are summarized in Table 1. In the following sections we consider some groups of interesting functionally related genes.

Signal transduction. Probably the most surprising feature of the EsV-1 genome is that it contains no fewer than

TABLE 1—*Continued*

49	79102	78746	118				
50	79165	80433	422	0.0006 0.0018 0.0097	86/360 (23%) 66/251 (26%) 15/55 (27%)	*Caenorhabditis elegans* H02F09.3 protein (length = 1275) [AF077538] Chlorovirus PBCV-1 hypothetical protein A14R (length = 1369) [U42580] Bacteriophage SPP1 product (fn3 domain) (length = 166) [X97918]	Bacterial regulatory proteins, gntR family signature (aa 240-264) [PS00043], Fibronectin type III domain (aa 3-71) [PF00041]
51	80481	80753	90				SP (aa 1-17), TM (aa 41-61, 70-90)
52	80854	81222	122				
53	81796	81236	186				TM (aa 85-105)
54	82162	81956	68				
55	82219	83196	325				PCC (aa 275-305)
56	83273	84919	548	9.7e-12 9.1e-09 6.8e-06	124/492 (25%) 106/459 (23%) 123/513 (23%)	*Rickettsia australis* outer membrane protein A (length = 2106) [AAD39531] *Yersinia pestis* YAPA protein (length = 1432) [AJ277624] *Caulobacter crescentus* S-LAYER protein (length = 1025) [P35828]	**EsV-1 calcium binding protein 1** [AF204951] EF-hand calcium-binding domain (aa 333-345) [PS00018], Thr-rich region (aa 378-420) [PS50325]
57	85024	86100	358	5.9e-52 7.7e-51	116/230 (50%) 116/227 (51%)	*Bos taurus* type II collogen (length = 347) [AF138883] *Equus caballus* type II collagen (length = 1418) [Q28396]	**Collagen-like protein 1** [AF204951] Collagen triple helix repeats (aa 104-162, 163-221, 251-310) [PF01391] [PS50288]
58	86203	87714	503	5.6e-06 4.5e-05	41/116 (35%) 53/185 (28%)	*Caenorhabditis elegans* W09G10.1protein (length = 222) [AF016671] *Triticum aestivum* glutenin (length = 766) [AF216868]	**vp55** [AF204951] Collagen triple helix repeat (aa 478-497) [PF01391] [PS50288], Gln-rich region (aa 396-453) [PS50322], MCC (aa 74-132), TM (aa 52-73), PCC (aa 467-501)
59	88136	87711	141	0.0015	26/102 (25%)	Chlorovirus PBCV-1 hypothetical protein A400R (length = 118) [U42580]	CC (aa 22-52)
60	88245	88715	156				SP (aa 1-35), TM (aa 53-73)
61	88765	89034	89	3.2e-05	19/58 (32%)	Chlorovirus PBCV-1 hypothetical protein A306L (length = 86) [U42580]	Pro-rich region (aa 6-26) [PS50099]
62	90368	89031	445	5.7e-16 1.2e-15	62/189 (32%) 53/170 (31%)	*Rattus norvegicus* dentin phosphoprotein precursor (length = 267) [U63111]. Human herpesvirus 8 latent nuclear antigen (length = 1036) [AF305694]	Glu-rich region (aa 31-175) [PS50313], Ser-rich region (aa 93-141) [PS50324], TM (aa 216-236), NCC (aa 116-175)
63	91031	90438	197				
64	91117	91743	208	4.0e-14 0.00041	55/202 (27%) 51/195 (26%)	Chlorovirus PBCV-1 hypothetical protein A166R (length = 268) [U42580] *Arabidopsis thaliana* F12A21.19 protein (length = 334) [AC008113]	
65	94056	91753	767	2.0e-49 3.1e-48 9e-48 2.6e-47	185/610 (30%) 140/338 (39%) 139/417 (33%) 142/369(8%)	*Azotobacter vinelandii* histidine protein kinase homolog GACS (length = 905) [AF197912] *Pseudomonas aeruginosa* two component sensor (length = 925) [AF030352] *Synechocystis* sp. sensory transduction histidine kinase slr2104 (length = 950) [S75136] *Synechocystis* sp. histidine kinase slr1759 (length = 1462) [D90903]	**Hybrid histidine kinase** Histidine kinase domain (aa 344-564) [PF00512] [PS50109] [HisKA] [HATPase_c], Response regulator receiver domain (aa 627-741) [PF00072] [PS50110] [REC], Prokaryotic membrane lipoprotein lipid attachment site (aa 10-20) [PS00013], SP (aa 1-19), TMs (aa 53-73, 299-319)
66	94110	95543	477	1.1e-48 2.1e-19	144/434 (33%) 105/427 (24%)	Chlorovirus PBCV-1 hypothetical protein A153R (length = 459) [U42580] Chilo iridescent virus (CIV) helicase (length = 459) [AF003534]	**Helicase** DEAD/DEAH box helicase (aa 120-211) [PF00270] [DEXDc], Helicases conserved C-terminal domain (aa 334-411) [PF00271] [HELICc]
67	95923	95549	124				SP (aa 1-33), TM (aa 11-31)
68	98036	95964	690				Gln-rich region (aa 78-90) [PS50322]
69	98543	98076	155	8.3e-07 1.2e-05	39/127 (30%) 44/140 (31%)	*Euplotes crassus* histone H1-1 (length = 152) [AF127331] *Drosophila hydei* histone H1 (length = 249) [X17072]	Lys-rich region (aa 67-132) [PS50318]
70	98715	100163	482				
71	100261	101223	320	1.7e-12 0.060	49/128 (38%) 61/246 (24%)	EsV-1-56 calcium-binding protein 1 (length = 548) [AAF28320] *Escherichia coli* 136.5 KDA lipoprotein (length = 1325) [AE000248]	Eukaryotic and viral aspartyl proteases active site (aa 165-176) [PS00141], Multispecific proteases of the proteasome (aa 125-183) [PS50247]
72	102065	101220	281				
73	102160	102405	81	3.3e-05 0.0012 0.0042	20/46 (43%) 17/46 (36%) 13/36 (36%)	*Strongyloides stercoralis* Igg-immunoreactive zinc finger protein (length = 211) [AF188206] *Phaseolus vulgaris* pathogenesis-related protein PVPR3 (length = 137) [M75856] Chlorovirus PBCV-1 hypothetical protein A623L (length = 67) [O41105]	AN1-like Zinc finger (aa 28-69) [PF01428] [ZnF_AN1]
74	102474	104003	509				NCC (aa 101-146), Ala-rich region (aa 213-249) [PS50310], Asp-rich region (aa 101-163) [PS50312], Gln-rich region (aa 7-96) [PS50322], Lys-rich

six genes for different hybrid histidine kinases, including the vhk-1 gene that we have previously described as a component of the viral envelope (Delaroque *et al.,* 2000b; Fig. 6).

Hybrid histidine kinases are members of a large protein family of two-component systems that serve as stimulus-response coupling mechanisms in bacteria, archaea, yeast, plants (*Arabidopsis,* tomato), and other eukaryotes (fungi, *Dictyostelium*). Their main function is to sense changes in the environment and to induce the appropriate genetic responses (Stock *et al.,* 2000). A hallmark of the hundreds of known two-component sig-

TABLE 1—*Continued*

							region (aa 181-238) [PS50318], Ser-rich region (aa 117-197) [PS50324], CC (aa 22-96)
75	104041	105222	393	9.4e-18 1.1e-17	75/251 (29%) 74/238 (31%)	*Ascaris suum* (pig roundworm) cathepsin B-like cysteine proteinase (length = 398) [U51892] *Caenorhabditis elegans* cathepsin B-like cysteine proteinase 3 precursor (length = 370) [L39890]	**Protease** Papain family cysteine protease (93-334) [PF00112], NCC (aa 263-292)
76	105297	106166	289	7.3e-15 2.2e-07	63/225 (28%) 24/54 (44%)	Chlorovirus PBCV-1 33K protein (length = 238) [U42580] *Arabidopsis thaliana* EN/SPM-like transposon protein (length = 140) [AC007232]	Asp-rich region (aa 233-283) [PS50312], Glu-rich region (aa 263-287) [PS50313], NCC (aa 252-287)
77	107170	106169	333				
78	107251	108126	291				
79	108534	108100	144				
80	109240	108596	214	3.8e-05 0.054	33/111 (29%) 52/195 (26%)	*Halobacterium cutirubrum* chaperone protein (length = 389) [AAB96891] *Synechocystis* sp. DnaJ protein (length = 377) [D64006]	DnaJ domain profile (aa 17-81) [PF00226] [PS50076] [DnaJ]
81	109298	109936	212				
82	109978	110892	304	5.5e-13 1.2e-12 1.3e-12 2.1e-12	68/246 (27%) 64/214 (29%) 68/213 (31%) 62/206 (30%)	*Arabidopsis thaliana* lectin-like protein kinase (length = 652) [AP002032] *Synechocystis* sp. serine/threonine protein kinase (length = 631) [D90915] *Arabidopsis thaliana* receptor-like protein kinase (length = 656) [AAB61102] *Arabidopsis thaliana* NPK1-related protein kinase 1S (length = 376) [AB000797]	**Protein kinase** Eukaryotic protein kinase domain (aa 16-179) [PF00069] [PS50011] [S_TKc], Protein kinases ATP-binding region signature (aa 22-43) [PS00107], Serine/Threonine protein kinases active-site signature (aa 128-140) [PS00108], Sugar transport proteins signature 2 (aa 31-56) [PS00217]
83	110936	111721	261	0.031 0.14 0.27 0.30	35/134 (26%) 34/130 (26%) 35/152 (23%) 38/139 (27%)	*Methanococcus jannaschii* UDP-glucose/GDP-mannose dehydrogenase (length = 427) [U67494] *Bacillus subtilis* GDP-mannose 6-dehydrogenase protein (length = 440) [Z92952] *Mycobacterium avium* GDHGA (length = 346) [AF143772] *Pseudomonas aeruginosa* UDP-glucose 6-dehydrogenase (length = 452) [AJ010734]	**UDP-glucose/GDP-mannose dehydrogenase** ATP/GTP-binding site motif A (aa 201-208) [PS00017], NAD binding site (aa 4-33) [PS50205], UBA NAD UBA/THIF-type NAD binding fold (aa 1-32) [PS50204], Prokaryotic membrane lipoprotein lipid attachment site (aa 74-84) [PS00013]
84	111792	113273	493	5.7e-21 2.3e-20 2.2e-16	105/350 (30%) 105/372 (28%) 93/331 (28%)	*Emericella nidulans* chitin synthase D (chsD) (length = 1852) [AB000125] *Blumeria graminis* chitin synthase (chs2) (length = 1867) [AF189366] *Aspergillus fumigatus* chitin synthase (chsE) (length = 1498) [Y09542]	**Glycosyltransferase** (EsV-1-CHS) [U95206] General glycosyltransferase domain (aa 75-225) [PS50167], SP (aa 1-31), TMs (aa 395-415, 422-442, 458-478) NCC (aa 205-235)
85	114427	113561	288	0.048	17/40 (42%)	Human immunodeficiency virus type 1 VPU protein (length = 81) [AF076998]	**TMB** [AF210454] TMs (aa 78-98, 120-140, 160-180, 183-203)
86	114840	114442	132				**TMA** [AF210454] SP (aa 1-22), TMs (aa 32-52, 67-87), MCC (aa 108-132)
87	115874	114900	324	2.0e-60 1.4e-59 2.3e-59	134/314 (42%) 133/319 (41%) 136/316 (43%)	*Arabidopsis thaliana* replication factor (length = 333) [AC010795] *Schizosaccharomyces pombe* RFC 37 kd subunit (length = 342) [AL035064] *Saccharomyces cerevisiae* RFC 37 kd subunit (length = 323) [L20502]	**RFC small subunit 1** [AF210454] ATPases associated with various cellular activities (aa 40-218) [PF00004] [AAA], ATP/GTP-binding site motif A (aa 45-52) [PS00017] [PS50101], Replication factor C conserved domain (aa 136-215) [PS50150]
88	117377	115992	461	5.7e-17 3.5e-15 2.5e-14	99/435 (23%) 99/437 (22%) 93/359 (26%)	*Schizosaccharomyces pombe* sensor-like histidine kinase (length = 2310) [Z98978] *Escherichia coli* sensor-regulator protein BARA (length = 918) [D10888] *Dictyostelium discoideum* hybrid histidine kinase DHKB (length = 1969) [AF024654]	**Hybrid histidine kinase (vhk-1)** [AF210454] Histidine kinase domain (aa 54-261) [PF00512] [PS50109] [HATPase_c], Response regulator receiver domain (aa 311-425) [PF00072] [PS50110] [REC]
89	117482	118354	290	1.2e-19	104/314 (33%)	*Drosophila melanogaster* CG7709 protein (length = 950) [AE003723]	Ala-rich region (aa 128-185) [PS50310], Pro-rich region (aa 43-118) [PS50099], Ser-rich region (aa 132-189) [PS50324]
90	118389	118709	106				
91	118751	120406	551	0.049	55/198 (27%)	Chlorovirus PBCV-1 hypothetical protein A533R (length = 374) [U42580]	
92	120466	120834	122	0.00077	22/78 (28%)	Human immunodeficiency virus type 1 GP120 (fragment) (length = 113) [M90866]	
93	120896	124036	1046	0.0	463/1016	*Feldmannia* sp. Virus DNA-dependent	**DNA polymerase**

nal transduction enzymes is their modular design (Grebe and Stock, 1999). The prototypic bacterial two-component system consists of two separate enzymes, a histidine protein kinase (HPK) and a response regulator. The histidine kinase has an aminoterminal domain for the reception of signals ("sensing domain") and a catalytic domain (HPK domain) with highly conserved amino acid sequence elements (homology boxes H, N, D, F, and G; Fig. 6). Upon activation, a conserved histidine in homology box H is autophosphorylated as a first step in signal

TABLE 1—*Continued*

					(45%)	DNA polymerase homolog (length = 996) [AF013260]	DNA polymerase family B (aa 135-834, 953-1012) [PF00136] [PS00116] [POLBc], Aldehyde dehydrogenases glutamic acid active site (aa 212-219) [PS00687]
				1.9e-99	289/903 (32%)	*Schizosaccharomyces pombe* DNA polymerase delta large chain (length = 1086) [X59278]	
				5.9e-94	255/786 (32%)	Chlorovirus NY-2A DNA polymerase (length = 913) [M86837]	
94	124090	124467	125				Leucine zipper pattern (aa 12-33) [PS00029], Myb DNA-binding domain repeat signature 1 (aa 50-58) [PS00037], TM (aa 70-90)
95	124671	124462	69				RecA family profile 2 (aa 10-69) [PS50163]
96	125923	125039	294	1.5e-08	36/81 (44%)	Chlorovirus PBCV-1 hypothetical protein A482R (length = 215) [U42580]	
97	126961	125984	325				Bacterial regulatory proteins, lysR family signature (aa 181-211) [PS00044], SP (aa 1-25)
98	127387	128061	224				
99	128306	128037	89				
100	128373	128774	133				
101	129720	128758	320	2.1e-31	76/197 (38%)	Chlorovirus PBCV-1 hypothetical protein A494R (length = 360) [U42580]	
				5.5e-05	55/227 (24%)	African swine fever virus B385R protein (length = 385) [U42580]	
102	129858	130157	99				
103	130205	131602	465	1.5e-13	70/252 (28%)	Chlorovirus PBCV-1 hypothetical protein A324L (length = 453) [U42580]	Pro-rich region (aa 80-118) [PS50099], Ser-rich region (aa 167-215) [PS50324], MCC (aa 213-285)
104	132521	131667	284	0.00073	43/169 (25%)	Human MAP/microtubule affinity regulating kinase (MARK) (length = 795) [AF154845]	**Protein kinase** Eukaryotic protein kinase domain (aa 75-193) [PF00069] [PS50011] [STYKc], Protein kinases ATP-binding region signature (aa 81-103) [PS00107]
				0.0023	39/142 (27%)	*Saccharomyces cerevisiae* cell cycle protein kinase CDC5/MSD2 (length = 705) [M84220]	
				0.0023	40/153 (26%)	*Caenorhabditis elegans* Ca2+/calmodulin-dependent protein kinase I (length = 348) [AB021864]	
105	132941	132525	138				
106	132954	133343	129				TM (aa 105-125)
107	134128	133847	93				TM (aa 39-59)
108	135011	134331	226				
109	136875	135055	606	1.0e-73	175/532 (33%)	Chlorovirus PBCV-1 hypothetical protein A456L (length = 654) [U42580]	Viral (Superfamily 1) RNA helicase (aa 562-577) [PF01443], ATP/GTP-binding site motif A (aa 340-347) [PS00017] [PS50101]
				1.1e-10	55/197 (27%)	*Haemophilus influenzae* phage phi-r73 primase-like protein (length = 589) [AF198256]	
				6.8e-10	87/325 (26%)	*Streptococcus thermophilus* bacteriophage SFi18 and SFi11 putative primase (length = 504) [AF158601]	
110	136965	137489	174				
111	137535	138878	447	9e-62	128/320 (40%)	*Feldmannia* sp. virus serine/threonine protein kinase (length = 418) [AF031820]	**Protein kinase** Eukaryotic protein kinase domain (aa 6-315) [PF00069] [PS50011] [S_TKc], Serine/Threonine protein kinases active-site signature (aa 125-137) [PS00108], Protein kinases ATP-binding region signature (aa 12-35) [PS00107], Prokaryotic membrane lipoprotein lipid attachment site (aa 248-258) [PS00013]
				2.2e-61	138/366 (37%)	*Arabidopsis thaliana* putative serine/threonine protein kinase (length = 441) [AC005623]	
				2.2e-61	138/366 (37%)	*Arabidopsis thaliana* CBL-interacting protein kinase 3 (CIPK3) (length = 375) [AF286051]	
				2.8e-61	119/264 (45%)	*Sorghum bicolor* serine/threonine kinase (SNFL1) (length = 440) [Y12464]	
112	140626	138875	583	9.1e-20	83/326 (25%)	*Vibrio cholerae* sensor histidine kinase/response regulator VC145 (length = 572) [AE004223]	**Hybrid histidine kinase** Histidine kinase domain (aa 235-444) [PF00512] [PS50109] [HisKA] [HATPase_c], Response regulator receiver domain (aa 469-580) [PF00072] [PS50110] [REC], Prokaryotic membrane lipoprotein lipid attachment site (aa 11-21) [PS00013], SP (aa 1-24), TMs (aa 41-61, 69-89)
				5.3e-19	94/384 (24%)	*Vibrio cholerae* sensor histidine kinase/response regulator VC1349 (length = 1331) [AE004214]	
				5.1e-18	91/333 (27%)	*Yersinia pestis* putative sensor protein EVGS (length = 803) [AL031866]	
				6.7e-18	96/367 (26%)	*Rhodobacter sphaeroides* DMSO/TMAO-sensor kinase (length = 815) [AF016236]	
113	140674	141000	108	0.0079	21/76 (27%)	*Rhodobacter sphaeroides* DMSR (fragment) (length = 170) [D89075]	**Phosphoshuttle** Histidine Phosphotransfer domain (aa 13-102) [HPT]
				0.043	27/92 (29%)	*Rhodobacter sphaeroides* DMSO/TMAO-sensor kinase (length = 815) [AF016236]	
114	141087	142181	364	0.043	34/124 (27%)	*Neurospora crassa* B3E4.110 protein (length = 387) [AL355931]	2 ankyrin repeats (aa 117-149, 188-217) [PF00023] [PS50297] [ANK]
115	144025	142439	528	1.7e-10	108/395 (27%)	*Arabidopsis thaliana* F22C12.10 protein (length = 646) [AC007764]	2Fe-2S ferredoxins, iron-sulfur binding region signature (aa 146-154) [PS00197]
				4.4e-08	79/271 (29%)	*Oryza sativa* P0501G01.5 protein (length =	

transduction. The second step is the transfer of the phosphoryl group from the histidine to an aspartate residue in the response regulator protein with its conserved receiver domain motifs (homology boxes 1, 2, and 3; Fig. 6). This then leads to a downstream effector that elicits the specific response (Hoch and Sihavy, 1995).

A variation of this molecular organization is the linkage of the histidine kinase and the receiver domain on one polypeptide chain known as a hybrid kinase. Hybrid kinases are rare in bacteria. For example, *Escherichia coli* expresses 30 different histidine kinases and 32 different response regulator proteins, but only five of these

TABLE 1—*Continued*

				0.0011	43/146 (29%)	646) [AP002819] Chlorovirus PBCV-1 hypothetical protein A154L (length = 347) [U42580]	
116	145269	146699	476	2.6e-38	124/429 (28%)	Chlorovirus PBCV-1 major capsid protein (A622L) (length = 437) [U42580]	**Major capsid protein**
117	147323	148897	524	1.1e-15 1.1e-11 5.3e-11 4.4e-05	59/200 (29%) 42/116 (36%) 56/170 (32%) 65/248 (26%)	*Lactococcus lactis* phage BK5-T hypothetical protein (length = 266) [L44593] *Lactobacillus casei* bacteriophage A2 putative antirepressor of the lysogenic cycle (length = 160) [AJ251789] *Melanoplus sanguinipes* entomopoxvirus MSV194 protein (length = 409) [AF063866] *Staphylococcus aureus* prophage phiPV83 antirepressor (length = 265) [AB044554]	**Antirepressor of the lysogenic cycle** BRO family, N-terminus (aa 1-101) [PF02498], Intron encoded nuclease repeat motif 1 (aa 316-369, 392-446, 471-523) [IENR1], Aminoacyl-transfer RNA synthetases class-II signature 1 (aa 261-277) [PS00179], CC (aa 110-146)
118	150524	149001	507	0.015	61/266 (22%)	*Rattus norvegicus* ribonuclease inhibitor (length = 456) [X62528]	
119	151881	150598	427	2.4e-10 6.6e-09	37/107 (34%) 81/312 (25%)	*Lactococcus* phage bIL170 endodeoxyribonuclease homolog e37 (length = 158) [AF009630] Chlorovirus PBCV-1 hypothetical protein A422R (length = 342) [U42580]	**Endonuclease** HNH endonuclease (aa 89-137, 275-324) [PF01844] [HNHc], Intron encoded nuclease repeat motif 1 (aa 148-203, 332-383) [IENR1]
120	153162	152719	147				TM (aa 24-44)
121	153522	153295	75				Lys-rich region (aa 13-32) [PS50318]
122	153960	154544	194				
123	154609	155052	147				TMs (aa 20-40, 113-133)
124	155492	155049	147				TMs (aa 13-33, 108-128)
125	156134	155574	186				
126	156769	156149	206	1.2e-06 0.01	63/189 (33%) 47/187 (25%)	*Deinococcus radiodurans* probable DNA polymerase III, epsilon subunit (length = 197) [AE001939] *Streptomyces coelicolor* putative DNA polymerase III, epsilon subunit (length = 328) [AL132644]	**Exonuclease** Exonuclease (aa 3-193) [PF00929] [EXOIII]
127	157102	156833	89				Leucine zipper pattern (aa 49-70) [PS00029]
128	158454	157195	419	6.4e-87 3.2e-85	174/342 (50%) 161/301 (53%)	*Trypanosoma brucei* ribonucleotide reductase (Class I, small subunit) (length = 337) [Y10768] *Cryptosporidium parvum* ribonucleotide reductase R2 subunit (length = 352) [AF275635]	**Ribonucleotide reductase (small subunit)** Ribonucleotide reductase (aa 132-399) [PF00268]
129	158633	160189	518	1.0e-07 7.0e-06 0.00011	39/101 (38%) 31/82 (37%) 32/70 (45%)	*Arabidopsis thaliana* T12J13.13 swib protein (length = 143) [AC009327] *Chlamydia trachomatis* probable swib (ym74) complex protein (length = 86) [AE001320] *Schizosaccharomyces pombe* SPCC285.17 swib protein (length = 233) [AL031545]	BAF60b domain of the SWIB complex (aa 394-480) [PF02201] [SWIB], Leucine zipper pattern (aa 232-253) [PS00029], Ala-rich region (aa 485-503) [PS500310]
130	160635	160198	145				
131	160929	160669	86				
132	161005	161910	301	1.4e-29 8.2e-23 8.5e-21	75/254 (29%) 61/254 (24%) 67/244 (27%)	Chlorovirus PBCV-1 PCNA homolog A193L (length = 262) [U42580] *Schizosaccharomyces pombe* PCN1 (length = 260) [X54857] *Saccharomyces cerevisiae* PCNA (length = 258) [X16676]	**Proliferating cell nuclear antigen** Proliferating cell nuclear antigen (aa 52-301) [PF00705]
133	162972	161911	353				TMs (aa 161-181, 197-217, 245-265, 271-291, 333-353)
134	163059	163640	193				Arg-rich region (aa 16-93) [PS500323], TM (aa 152-172)
135	165389	163746	547	4.6e-26 5.8e-26	72/221 (32%) 88/267 (32%)	*Saccharomyces cerevisiae* suppressor protein SRP40 (length = 406) [L11275] *Staphylococcus aureus* clumping factor B precursor (length = 913) [AJ224764]	Asp-rich region (aa 349-442) [PS500312], Glu-rich region (aa 364-542) [PS500313], Pro-rich region (aa 295-333) [PS500099], Ser-rich regions (aa 177-202, 420-543) [PS500324], NCCs (aa 349-419, 458-526)
136	165820	165464	118				TMs (aa 61-81, 98-118)
137	165925	166356	143	0.00023	29/124 (23%)	Chlorovirus PBCV-1 A470 R (length = 203) [U42580]	TM (aa 74-94)
138	166394	167602	402	0.00025 0.00080 0.00092	51/220 (23%) 50/235 (21%) 30/126 (23%)	*Pyrococcus horikoshii* RFC, large subunit (length = 468) [AP000001] *Methanococcus jannaschii* RFC, large subunit (length = 516) [U67532] *Sulfolobus solfataricus* replication factor C, large subunit (length = 405) [AL512964]	**RFC large subunit** ATPases associated with various cellular activities (aa 66-328) [PF00004] [AAA], ATP/GTP-binding site motif A (aa 71-78) [PS00017] [PS50101]
139	168423	167599	274	1.1e-27	67/210 (32%)	*Drosophila melanogaster* oligoribonuclease (length = 211) [AE003744]	**Oligoribonuclease** Exonuclease (aa 44-267) [PF00929] [EXOIII]

are hybrid kinases (Mizuno, 1997). In contrast, hybrid kinases are the common form of two-component systems in eukaryotes (Stock *et al.,* 2000). In agreement with this, the putative six two-component enzymes encoded by the EsV-1 genome are hybrid kinases. Five of the six enzymes contain long aminoterminal domains for the reception of stimuli, including possibly light stimuli, since the hybrid kinase encoded by ORF 181 has an aminoterminal extension with high similarity to the phytochrome chromophore-binding domain found in corresponding enzymes of plants, the nonphotosynthetic bacterium *Deinococcus*, and the photosynthetic cyanobacterium

TABLE 1—*Continued*

				3.8e-27	65/172 (38%)	Human oligoribonuclease, mitochondrial precursor (length = 237) [AL110239]	
				4.8e-27	74/194 (38%)	*Caenorhabditis elegans* probable oligoribonuclease (length = 193) [Z72502]	
140	168542	169369	275	0.26	21/53 (39%)	Chlorovirus PBCV-1 hypothetical protein A225L (length = 90) [U42580]	CC (aa 182-206)
141	170004	169378	208	2.4e-16	44/130 (33%)	Chlorovirus PBCV-1 hypothetical protein A471R (length = 173) [U42580]	
142	170007	171344	445	1.5e-15	65/219 (29%)	*Arabidopsis thaliana* ankyrin-like protein F7O18.18 (length = 456) [AC011437]	**Ring finger protein 1** [AF204952] Ring finger, C3H2C3 type (RING finger) (aa 87-127) [PF00097] [PS50089] [RING], 5 ankyrin repeats (aa 255-285, 289-321, 324-354, 355-387, 388-420) [PF00023] [PS50297] [ANK], Gln-rich region (aa 13-27) [PS50322], Asp-rich region (aa 137-161) [PS50312], NCC (aa 137-169)
				1.3e-11	76/250 (30%)	*Arabidopsis thaliana* ankyrin repeat protein EMB506 (length = 315) [AB010699]	
				9.4e-08	25/97 (25%)	*Arabidopsis thaliana* ring finger protein F4I1.14 (length = 180) [AC004521]	
				8.8e-05	41/123 (33%)	Cercopithecine herpesvirus 7 transactivator (length = 503) [AF275348]	
143	171389	172144	251				**vp27** [AF204952] Ser-rich region (aa 45-57) [PS50324]
144	172311	174407	698	1.1e-26	50/72 (69%)	*Lycopersicon esculentum* TFM5 protein (length = 207) [X95262]	**vp74** [AF204952] Gly-rich regions (aa 52-63, aa 440-510) [PS50315], Ala-rich region (aa 337-425) [PS50310], CC (aa 103-439)
				4.3e-25	127/441 (28%)	Human keratin type I cytoskeletal 9 (length = 622) [X75015]	
145	174443	175360	305				TMs (aa 20-40, 59-79, 150-170, 184-204, 255-275)
146	175594	174677	305	2.4e-05	27/81 (33%)	*Neurospora crassa* conserved hypothetical protein B1D1.390 (length = 805) [AL355927]	Prokaryotic membrane lipoprotein lipid attachment site (aa 220-230) [PS00013], Ser-rich region (aa 213-240) [PS50324], Gly-rich region (aa 214-238) [PS50315]
147	175744	177957	737				Bacterial regulatory proteins, lysR family signature (aa 79-109) [PS00044], Lipoxygenase homology 2 region (aa 36-48) [PS50095], NCC (aa 253-283)
148	179816	177981	611				Glu-rich region (aa 171-221) [PS50313], PCC (aa 243-270), NCC (aa 186-206), MCC (185-221), CCs (aa 27-55, 153-230)
149	179752	180714	320				
150	180748	181347	199				OTU domain (aa 6-169) [PS50802], TM (aa 113-133)
151	182813	181362	483				Ala-rich region (aa 312-385) [PS50310], Lys-rich region (aa 305-351) [PS50318], CC (aa 360-406)
152	182915	183364	149				TonB-dependent receptor proteins signature 1 (aa 1-58) [PS00430], TM (aa 94-114)
153	183432	183758	108				Lys-rich region (aa 13-98) [PS50318]
154	185167	183779	462	2.0e-05	58/184 (31%)	*Schizosaccharomyces pombe* putative DNA repair protein, yeast rad50 homolog (length = 666) [AL360094]	K-box region (aa 316-412) [PF01486], Leucine zipper pattern (aa 336-357) [PS00029], CC (aa 249-274, 324-417)
				2.4e-05	60/211 (28%)	*Streptococcus pyogenes* M5 8193 protein (length = 457) [U02480]	
155	197924	196773	383	0.0	382/383 (99%)	EsV-1-170	**Transposase** Transposase (IS4 family) (aa 51-282) [PF01609]
				1.5e-06	65/234 (27%)	*Lactococcus lactis* putative transposase (length = 439) [U91581]	
				0.025	67/285 (23%)	*Enterobacter agglomerans* transposase (length = 439) [X81894]	
156	198119	199072	317	1.0	30/111 (27%)	Vaccinia virus 30 kDa protein kinase homolog (length = 300) [P20505]	**Protein kinase** Eukaryotic protein kinase domain (aa 68-317) [PF00069] [PS50011] [STYKc]
				1.0	30/111 (27%)	Variola virus protein kinase homolog B1R (length = 300) [L22579]	
157	200281	201219	312	3.9e-13	55/161 (34%)	Fowlpox virus ORF FPV023 ankyrin repeat gene family protein (length = 434) [AF198100]	4 ankyrin repeats (aa 104-134, 135-164, 165-197, 198-233) [PF00023] [PS50297] [ANK]
				1.4e-12	65/178 (36%)	*Arabidopsis thaliana* putative ankyrin T4M8.14 (length = 247) [AC006284]	
158	201933	201313	206	0.12	33/100 (33%)	*Drosophila melanogaster* CG4565 protein (length = 223) [AE003688]	SET-domain of transcriptional regulators (TRX, EZ, ASH1, Su(var)3-9..) (aa 51-187) [PF00856] [PS50280] [SET]
				0.16	33/137 (24%)	Human pr/set domain containing protein 07 (length = 345) [AF287261]	
159	202176	203741	521				Fibronectin type III domain (aa 290-379) [PF00041] [FN3]
160	207366	203743	1207	3.0e-09	77/254 (30%)	*Arabidopsis thaliana* glycine-rich protein F7K2.60 (length = 608) [AL033545]	Prokaryotic membrane lipoprotein lipid attachment site (aa 306-316) [PS00013] Trp-Asp (WD) repeats signature (aa 893-907) [PS00678], TMs (aa 90-110, 180-200, 209-229, 304-324, 355-375, 385-405, 489-509, 642-662, 736-756)
				5.7e-07	136/621 (21%)	*Caenorhabditis elegans* H02F09.3 protein (length = 1275) [AF077538]	
				4.2e-05	160/707 (22%)	Chlorovirus PBCV-1 hypothetical protein A540L (length = 1176) [U42580]	
161	209443	209985	180	4.0e-07	30/108 (27%)	Chlorovirus PBCV-1 ERV1 protein	TMs (aa 11-31, 160-180)

Synechocystis (Davis *et al.,* 1999; Kotani and Tabata, 1998; Fig. 6). In *Synechocystis*, the photoreceptor is composed of an amino terminal domain where the tetrapyrrole chromophore is covalently linked to a conserved cysteine residue and a carboxyl-terminal HPK domain (Yeh *et al.,* 1997). This is related to the molecular design of the EsV-1 ORF 181-encoded polypeptide (Fig. 6A). However, in contrast to the *Synechocystis* protein, the putative EsV-1 protein does not contain a cysteine at the right position in the conserved region. It is therefore likely that the chromophore in the EsV-1 enzyme is attached to an adjacent histidine residue (Fig. 6B) exactly

TABLE 1—*Continued*

				1.5e-05	28/91 (30%)	homolog A465R (length = 118) [U42580] *Saccharomyces cerevisiae* ERV1 protein precursor (length = 189) [X60722]	
162	210648	210001	215				Pro-rich region (aa 126-199) [PS50099], Gln-rich region (aa 38-87) [PS50322], Lys-rich region (aa 6-31) [PS50318], Tyr-rich region (aa 46-100) [PS50328], PCC (aa 6-32)
163	211503	210724	259				Leucine zipper pattern (aa 11-32), SP (aa 1-30)
164	213702	211615	695	6e-43 0.00022 0.00062	115/321 (35%) 53/169 (31%) 53/170 (31%)	*Streptomyces coelicolor* putative secreted protein (length = 358) [AL133236] *Bradyrhizobium japonicum* nosD product (length = 453) [AJ002531] *Achromobacter cycloclastes* nosD product (length = 442) [Y15161]	ATP/GTP-binding site motif A (aa 200-207) [PS00017], Discoidin domain (FA5/8 type C domain) (aa 417-534, 555-695) [PS50022], Glu-rich region (aa 522-640) [PS50313], MCC (aa 537-562), SP (aa 1-16)
165	214169	213855	104				
166	214596	215723	375	2.9e-07 2.5e-06 6.1e-06	35/103 (33%) 52/172 (30%) 42/134 (31%)	*Aspergillus aculeatus* FII-CMCASE (length = 412) [AB015510] *Drosophila melanogaster* CG15786 protein (length = 283) [AE003433] *Aspergillus aculeatus* exoglucanase I precursor (length = 540) [AB002821]	Fungal cellulose binding domain (aa 341-370) [PF00734] [fCBD], Thr-rich region (aa 296-331) [PS50325], SP (aa 1-16)
167	216556	215732	274				MCC (aa 119-159)
168	216619	217119	166	0.081 1.2	26/78 (33%) 22/89 (24%)	*Salmonella typhi* putative partition protein (length = 171) [AF250878] *Campylobacter jejuni* putative secreted nuclease (length = 175) [AL139076]	**Nuclease** Staphylococcal nuclease homologues (aa 15-165) [PF00565] [SNc]
169	217194	218375	393	5.3e-11 2.1e-07	70/175 (40%) 49/158 (31%)	*Arabidopsis thaliana* thaumatin-like protein YPR5 (length = 301) [AL022373] *Arabidopsis thaliana* osmotin-like protein osm34 precursor (length = 244) [X89008]	**Thaumatin-like protein** Thaumatin (aa 159-194, 352-383) [PF00314] [THN], SP (aa 1-22)
170	223541	222390	383	0.0 1.5e-06 0.025	382/383 (99%) 65/234 (27%) 67/285 (23%)	EsV-1-155 *Lactococcus lactis* putative transposase (length = 439) [U91581] *Enterobacter agglomerans* transposase (length = 439) [X81894]	**Transposase** Transposase (IS4 family) (aa 51-282) [PF01609]
171	232829	236530	1233	7.9e-49 3.3e-44	294/1124 (26%) 144/519 (27%)	*Trypanosoma cruzi* R27-2 protein (length = 1128) [L04603] *Babesia bigemina* 200 KDA antigen P200 (length = 1108) [AF142406]	Leucine zipper pattern (aa 203-224) [PS00029], cdc15 (*S.pombe*) N-terminal domain (aa 856-1085) [PS50133], Ala-rich region (aa 380-617) [PS50310], Lys-rich region (aa 116-621) [PS50318], Glu-rich regions (aa 680-689, 899-1107) [PS50313], NCC (aa 680-699), CCs (aa 113-144, 205-247, 329-438, 494-631, 671-699, 828-1090), SP (aa 1-19)
172	239346	236917	809	3e-59	199/721 (27%)	*Feldmannia* sp. virus ORF2 (length = 673) [U22375]	Ring finger, C3HC4 type (RING finger) (aa 567-609) [PF00097] [PS50089] [PS00518] [RING], ATP/GTP-binding site motif A (by similarity) (aa 241-248), Sir2 family conserved domain (aa 429-596) [PS50305]
173	239739	240830	363	2.4e-55 5.0e-55 8.2e-55	130/363 (35%) 132/364 (36%) 129/373 (34%)	*Solanum tuberosum* dnaJ protein homolog (length = 419) [X94301] *Cucumis sativus* dnaJ protein homolog (length = 413) [X67695] *Dictyostelium discoideum* heat shock protein DDJ1 (length = 413) [AF063011]	DnaJ domain (aa 8-69) [PF00226] [PS50076] [PS00636] [DnaJ], DnaJ central domain (4 repeats) (aa 143-225) [PF00684], DnaJ C terminal region (aa 238-359) [PF01556]
174	240865	241224	119				TMs (aa 39-59, 74-94)
175	242354	243406	350	6.6e-05	38/153 (24%)	*Borrelia burgdorferi* plasmid cp26 hypothetical 54.3 kd protein (length = 449) [AE000792]	
176	243546	244310	254	4.8e-18 1.2e-16	61/182 (33%) 60/182 (32%)	*Synechocystis* sp. hypothetical slr0645 22.8 kD protein (length = 206) [D64002] *Synechocystis* sp. hypothetical 22.4 kD protein (length = 200) [D90912]	Von Willebrandt factor type A domain (aa 59-247) [PF00092] [PS50234] [VWA]
177	244988	244785	67				TM (aa 47-67)
178	245182	247263	693				
179	248286	247603	227				
180	248410	250707	765	1.2e-234 1.8e-233	431/768 (56%) 429/768 (55%)	*Mus musculus* ribonucleoside-diphosphate reductase chain M1 (length = 792, large subunit) [K02927] Human ribonucleoside-diphosphate reductase chain M1 (length = 792, large subunit) [X59543]	**Ribonucleotide reductase (large subunit)** Ribonucleotide reductase (aa 127-732) [PF00317], ROM/Citron/NIK putative rho binding motif (aa 268-304) [PS50219]
181	253826	251025	933	1.9e-25 9.7e-20	135/553 (24%) 130/460 (28%)	*Pseudomonas aeruginosa* probable bacteriophytochrome (length = 728) [AE004828] *Deinococcus radiodurans* photoreceptor (length = 755) [AE001862]	**Hybrid histidine kinase** GAF domain (aa 150-318) [PF01590] [GAF], Histidine kinase domain (aa 535-777) [PF00512] [PS50109] [HisKA]

as described for the photoreceptor from *Deinoccocus* (Davis *et al.,* 1999).

In addition, the EsV-1 genome encodes a small protein related to known histidine-containing phosphotransfer (HPt) domains (Fig. 6). In bacteria, HPt domains are usually modules of hybrid kinases, but in eukaryotes HPts are found as separate proteins serving as stations in phosphoryl-transfer relay systems. For example, the genome of *Arabidopsis* contains three different genes for HPts (Miyata *et al.,* 1998; Suzuki *et al.,* 1998).

TABLE 1—*Continued*

				3.2e-18	151/608 (24%)	*Anabaena* sp cyanobacterial phytochrome A (length = 765) [AB028873]	[HATPase_c],
				5.0e-16	114/445 (25%)	*Arabidopsis thaliana* phytochrome E (length = 1112) [X76610]	Response regulator receiver domain (aa 798-924) [PF00072] [PS50110] [REC],
				1.9e-15	122/423 (28%)	*Synechocystis* sp. hypothetical 84.2 kd photoreceptor SLR0473 (length = 748) [D64001]	Phytochrome chromophore attachment site domain profile (aa 150-308) [PS50046]
				2e-16	163/712 (22%)	*Arabidopsis thaliana* phytochrome C (length = 1111) [P14714]	
182	253899	254879	326	8.3e-78	155/322 (48%)	*Oryza sativa* replication factor C 37 kDa subunit (length = 339) [AB045677]	**RFC small subunit**
				2.7e-73	160/330 (48%)	*Arabidopsis thaliana* F8K7.11 protein (length = 319) [AC007727]	Viral (Superfamily 1) RNA helicase (aa 38-113) [PF01443] [DEXDc],
				5.0e-71	154/326 (47%)	Human activator 1 37 kd subunit (replication factor C 37 kd subunit) (length = 363) [M87339]	DEAD/DEAH box helicase (aa 106-116) [PF00270], ATPases associated with various cellular activities (aa 37-215) [PF00004] [AAA], ATP/GTP-binding site motif A (aa 42-49) [PS00017] [PS50101],
				1.9e-69	150/330 (45%)	*Saccharomyces cerevisiae* activator 1 41 kDa subunit (length = 353) [D28499]	Replication factor C conserved domain (aa 135-217) [PS50150]
183	254986	255393	135	3.0e-09	39/132 (29%)	*Feldmannia* sp. virus ORF3 (length = 137) [U22375]	
184	255485	256198	237				
185	256326	257099	257	1.2e-08	58/216 (26%)	*Aspergillus niger* ferulic acid esterase A (length = 281) [Y09330]	**Lipase**
				9.3e-08	41/161 (25%)	*Synechocystis* sp SLL0482 protein (length = 407) [D64004	Lipase (class 3) (aa 69-247) [PF01764], Ser-rich region (aa 10-18) [PS50324], SP (aa 1-19)
				9.5e-08	53/178 (29%)	*Thermomyces lanuginosus* lipase (length = 291) [AF054513]	
186	257129	259171	680	1.0e-21	107/407 (26%)	*Pseudomonas fluorescens* sensor kinase (length = 575) [L29642]	**Hybrid histidine kinase**
				1.7e-20	102/370 (27%)	*Deinococcus radiodurans* sensor histidine kinase/response regulator (length = 577) [AE001826]	Histidine kinase domain (aa 323-535) [PF00512] [PS50109] [HATPase_c] Response regulator receiver domain (aa 554-667) [PF00072] [PS50110] [REC],
				3.8e-20	112/412 (27%)	*Vibrio cholerae* sensor histidine kinase/response regulator (length = 1331) [AE004214]	MCC (aa 403-430), TMs (aa 29-49, 76-96, 116-136, 158-178, 185-205, 208-228, 232-252, 273-293)
187	259449	260423	324	4.5e-47	121/331 (36%)	Human replication factor C 38K chain (length = 356) [L07541]	**RFC small subunit**
				7.7e-43	118/316 (37%)	*Caenorhabditis elegans* C39E9.13 protein (length = 354) [Z70307]	Replication factor C conserved domain (aa 146-257) [PS50150],
				1.3e-42	108/331 (32%)	*Drosophila melanogaster* LD06837 protein (length = 356) [AE003631]	ATPases associated with various cellular activities (aa 20-216) [PF00004] [AAA],
				1.6e-33	87/262 (33%)	*Saccharomyces cerevisiae* activator 1 subunit 5 (length = 354) [U26031]	ATP/GTP-binding site motif A (aa 25-32) (by similarity)
188	260445	261758	437				Prokaryotic membrane lipoprotein lipid attachment site (aa 35-45) [PS00013], Prenyl group binding site (CAAX box) (aa 434-437) [PS00294], TMs (aa 32-52, 59-79, 110-130, 133-153, 166-186, 200-220, 253-273, 276-296, 317-337, 351-371), MCC (aa 234-253)
189	261799	262641	280	4.0e-05	37/115 (32%)	*Streptomyces phaeochromogenes* plasmid pJV1 SPDB2 protein (length = 471) [Q54674]	MCC (aa 174-246), CC (aa 208-255), Arg-rich region (aa 167-242) [PS50323], Glu-rich region (aa 199-237) [PS50313]
190	262786	263298	170				Lipocalin signature (aa 49-62) [PS00213], SP (aa 1-19)
191	264683	263295	462				Cys-rich region (aa 429-448) [PS50311], MCC (aa 88-121), CC (aa 103-131)
192	267902	265119	927				
193	268478	268317	53	0.00047	16/34 (47%)	*Drosophila melanogaster* CG10756 protein (TFIID-18) (length = 136) [AE003665]	Transcription initiation factor IID, 18kD subunit (aa 10-49) [PF02269],
				0.00060	15/34 (44%)	Human TAFII18 (length = 124) [X84003]	TM (aa 33-53)
194	269261	268845	138				
195	270177	269740	145				Cys-rich region (aa 31-40) [PS50311], TM (aa 21-41)
196	270169	270396	75	3.2e-05	19/38 (50%)	*Drosophila melanogaster* CG10756 protein (TFIID-18) (length = 136) [AE003665]	Transcription initiation factor IID, 18kD subunit (aa 33-70) [PF02269]
				4.1e-05	18/38 (47%)	Human TAFII18 (length = 124) [X84003]	
				0.00077	12/39 (30%)	*Schizosaccharomyces pombe* putative TFIID 19 kd subunit (length = 132) [AL023776]	
197	270835	272211	458	34e-5	70/299 (23%)	Bacteriophage phi-C31 repressor protein C (length = 683) [X12865]	**Repressor of the lysogenic cycle**
198	274830	272218	870				
199	275146	276396	416	2.4e-27	79/231 (34%)	Human ankyrin G product (length = 4377) [U13616]	7 ankyrin repeats (aa 11-41, 45-77, 78-110, 113-145, 146-178, 179-208, 213-245) [PF00023] [PS50297] [ANK]
				3.9e-27	79/231 (34%)	Human ankyrin G119 product (length = 1088) [U43965]	

The functions of the hybrid kinases in the EsV-1 infection cycle are not known, but they could be somehow involved in the regulation of the viral latency, as the related *Chlorella* virus PBCV-1 with its lytic infection cycle has no genes for histidine kinase-like enzymes. PBCV-1 encodes instead six different Tyr- and Ser/Thr-specific protein kinases while the EsV-1 genome contains only four genes which could encode Ser/Thr-specific protein kinases of which one (ORF 111) has been previously described for another brown algal virus (FsV) (Lee *et al.,* 1998a).

TABLE 1—*Continued*

200	276468	277250	260				
201	278968	279177	69				TM (aa 27-47)
202	279200	279400	66				
203	280953	280651	100				
204	282526	281174	450				Ala-rich regions (aa 17-63, 421-443) [PS50310], MCC (aa 261-323), CC (aa 254-286), TMs (aa 131-151, 154-174, 340-360)
205	285462	284833	209				
206	286008	285565	147				Prokaryotic membrane lipoprotein lipid attachment site (aa 7-17) [PS00013]
207	293009	293626	205				
208	293938	294990	350	1.0e -107 0.0018	178/330 (53%) 21/71 (29%)	EsV-1-20 Human hypothetical 15.8 kDa protein (length = 136) [AL137671]	Ring finger, C3H2C3 type (RING finger) (aa 145-192) [PF00097] [PS50089] [RING], MCC (aa 233-312), SP (aa 1-19)
209	296371	295172	399				
210	296503	298032	509				
211	298485	299981	498				Pro-rich region (aa 116-123) [PS50099]
212	300345	300088	85				
213	300736	303096	786				**Integrase** "Phage" integrase family (aa 483-686) [PF00589]
214	309229	308969	86				ATP/GTP-binding site motif A (aa 46-53) [PS00017], TM (aa 64-84)
215	311656	311901	81				Prokaryotic membrane lipoprotein lipid attachment site (aa 53-63) [PS00013], TM (aa 50-70)
216	313449	312700	249	1.0e-08 9.9e-06	57/204 (27%) 45/211 (21%)	*Aeropyrum pernix* hypothetical protein APE1216 (length = 533) [AP000061] *Halobacterium halobium* sensory rhodopsin I transducer (length = 535) [L05603]	CC (aa 50-76)
217	313719	314450	243				Arg-rich region (aa 10-102) [PS50323], MCC (aa 45-105), CC (aa 42-104)
218	314586	315425	279				Ala-rich region (aa 123-261) [PS50310], Lys-rich region (aa 49-254) [PS50318], CC (aa 131-179)
219	316880	317113	77	0.00029 0.041	21/50 (42%) 22/63 (34%)	human hypothetical 12.5 kda protein (length = 116) [AL512683] *Mus musculus* nuclear localization signal binding protein (length = 143) [S79410]	His-rich region (aa 2-21) [PS50316], TM (aa 15-35)
220	318190	318456	88				TM (aa 40-60)
221	318549	318758	69				
222	320964	318883	693				
223	321878	321504	124	1.9e-09 3.4e-07 1.9e-05	32/77 (41%) 27/97 (27%) 24/72 (33%)	Chlorovirus PBCV-1 K^+ channel protein A250R (length = 94) [U42580] *Methanococcus jannaschii* MJ0139 protein (length = 209) [U67471] Human cyclic GMP gated potassium channel (length = 511) [U96110]	**Potassium channel** Pore region of potassium channels (aa 72-124) [PS50265], Leucine zipper pattern (aa 101-122) [PS00029], TMs (aa 36-56, 68-88, 99-119)
224	321942	322916	324	1.1e-48 1.7e-47 4.5e-47 3.1e-46	107/322 (33%) 112/317 (35%) 113/320 (35%) 108/320 (33%)	*Plasmodium falciparum* replication factor C3 (length = 344) [AF069296] *Arxula adeninivorans* replication factor C subunit (length = 344) [AJ007712] Human activator 1 36 kd subunit (length = 340) [L07540] *Saccharomyces cerevisiae* activator 1 40 kDa subunit (length = 340) [L18755]	**RFC small subunit** Viral (Superfamily 1) RNA helicase (aa 38-52) [PF01443], DEAD/DEAH box helicase (aa 98-113) [PF00270] [DEXDc], ATPases associated with various cellular activities (aa 37-215) [PF00004] [AAA], Replication factor C conserved domain (aa 134-213) [PS50150], ATP/GTP-binding site motif A (aa 42-49) [PS00017] [PS50101]
225	322971	323228	85	0.090	19/72 (26%)	*Bacillus halodurans* RECR (DNA repair and genetic recombination) (length = 198) [AP001507]	
226	325898	327883	661	6e-20 8e-15	79/292 (27%) 72/257 (28%)	*Azotobacter vinelandii* mannuronan C-5-epimerase (length = 525) [X87973] *Pseudomonas aeruginosa* mannuronan C-5-epimerase (length = 543) [U27829]	**EsV-1 gp1** [S49901] MCC (aa 580-623)
227	328326	328943	205				CC (aa 81-109)
228	329019	329276	85				Lys-rich region (aa 57-73) [PS00318]
229	330162	329413	249				Arg-rich region (aa 17-31) [PS00323]
230	330491	330721	76				
231	333659	332508	383				

Thus, *Chlorella* viruses and phaeoviruses have the potential to build up complex, but different, phosphate transfer systems which could modify viral and host protein functions, but also transmit external and internal stimuli. This may also be the function of yet another protein encoded by the genomes of both viruses, a potassium channel component (Plugge *et al.,* 2000) (EsV-1 ORF 223). In any case, the presence of two types of signal-transducing protein kinases in EsV-1 is unprecedented for a virus.

DNA metabolism and DNA replication. To guarantee the supply of deoxynucleotides in nonproliferating host cells, large DNA viruses frequently encode enzymes with functions in deoxynucleotide synthesis. For example, the PBCV-1 genome contains more than a dozen genes which could be involved in nucleotide metabolism (Van Etten and Meints, 1999). In contrast, EsV-1 seems to encode only an ATPase (ORF 26) as well as the small subunit (ORF 128) and the large subunit (ORF 180) of

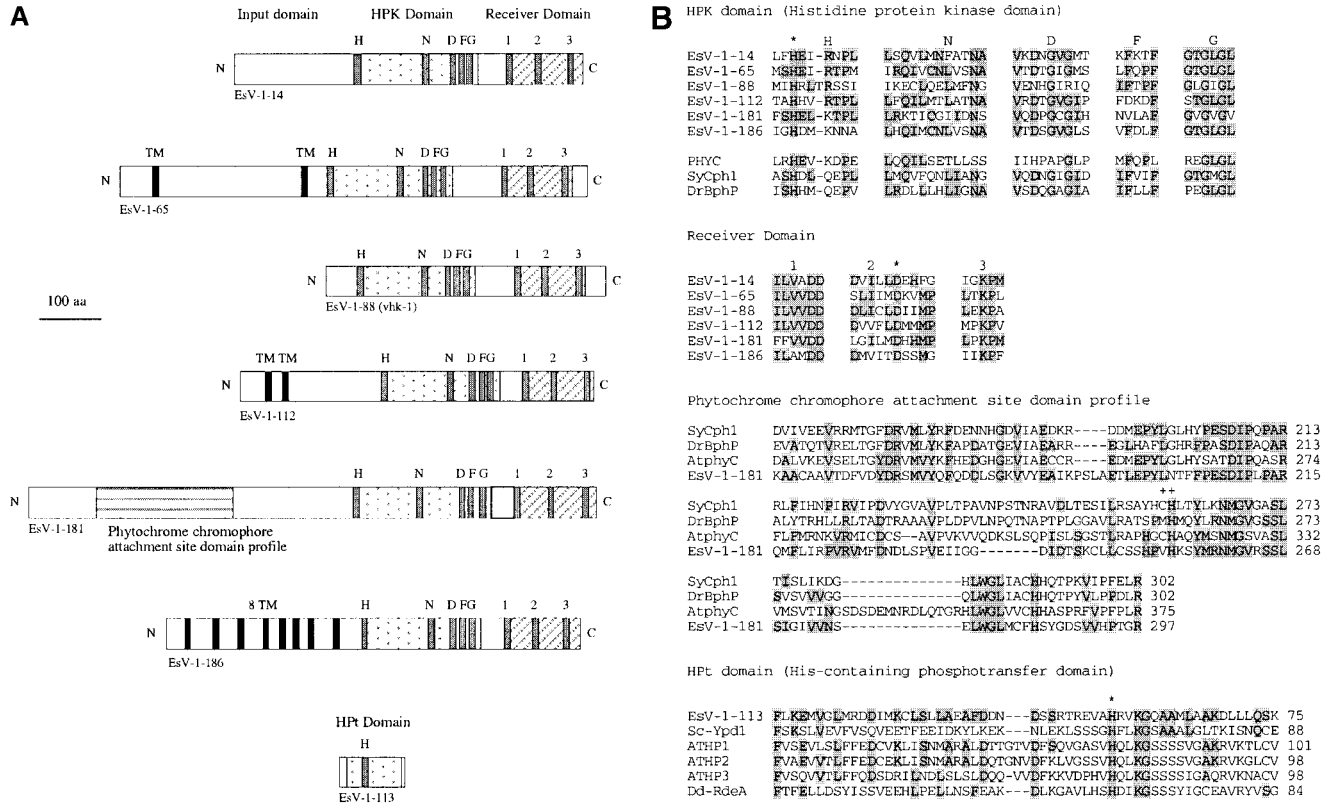

FIG. 6. The EsV-1 two-component system. (A) Schematic diagram of the viral hybrid kinases and the phosphoshuttle component. The histidine protein kinase (HPK) domain (stippled) contains the homology boxes H, N, D, F, and G. The receiver domains (hatched) include the homology boxes 1, 2, and 3. HPt, His-containing phosphotransfer; TM, hydrophobic transmembrane domains; N, amino terminal ends; C, carboxyl-terminal ends. (B) Top: comparison of the homology boxes in the viral histidine protein kinases and photoreceptors. Note that phytochromes do not contain a receiver domain. Bottom: the viral chromophore binding domain and His-containing phosphotransfer (HPt) are compared with corresponding proteins. *Arabidopsis thaliana* phytochrome C (PHYC, GenBank Accession No. P14714), *Synechocystis* sp. Cph1 (SyCph1, GenBank Accession No. Q55168), *Deinococcus radiodurans* BphP (DrBphP, GenBank Accession No. Q9RZA4), *Saccharomyces cerevisiae* Ypd1 (Sc-Ypd1, GenBank Accession No. NP_010046), *Arabidopsis thaliana* HP1, 2, and 3 (ATHP1, 2, 3, GenBank Accession Nos.: BAA37110, BAA37111, and BAA37112), *Dictyostelium discoideum* RdeA (Dd-RdeA, GenBank Accession No. AAC61850). Asterisks, phosphoaccepting amino acids. +, cysteine and histidine to which the chromophore is ligated in the plant and bacteria counterparts.

ribonucleotide reductase, a key enzyme in deoxynucleotide synthesis.

Deoxynucleotides are the substrates for the EsV-1-encoded putative DNA polymerase (ORF 93). Its predicted amino acid sequence is quite similar to those of DNA polymerases encoded by other phycodnaviruses (FsV) (Lee *et al.,* 1998b; PBCV-1, Grabherr *et al.,* 1992). These enzymes belong to the eukaryotic DNA-polymerase-δ family (Villarreal and DeFilippis, 2000) and contain a domain for a proofreading 3′-5′-exonuclease (Hübscher *et al.,* 2000). Curiously, EsV-1 seems to possess an additional gene with the potential to code for a proofreading exonuclease (ORF 126), and this putative exonuclease shares 30% amino acid identity with the ϵ-subunit of bacterial DNA polymerase III.

We have also identified an EsV-1 gene (ORF 132) for PCNA (proliferating cell nuclear antigen), the sliding clamp processivity factor, and, more surprisingly, a series of five genes (ORF 87, ORF 138, ORF 182, ORF 187, and ORF 224) that together could encode the heteropentameric replication factor C (RFC) required for the ATP-dependent loading of PCNA (Mossi and Hübscher, 1998). The interesting point here is that the *Chlorella* virus PBCV-1 also has genes for DNA polymerase δ and for PCNA, but not for RFC, but instead codes for other essential replication functions such as a DNA ligase and a DNA topoisomerase II (Van Etten and Meints, 1999; Lavrukhin *et al.,* 2000). Thus, each virus genome contains genes for essential elements of the eukaryotic replication machine, but none has the full complement of replicative genes. In particular, both viral genomes lack genes for a eukaryotic primase function, but possess a gene (ORF 109 in EsV-1; A456L in PBVC-1; Kutish *et al.,* 1996) with some sequence similarities to bacteriophage-encoded primase–helicases. It will be of great interest to determine whether this bacteriophage-like enzyme would provide the essential primase function required for phycodnaviral DNA replication because this would be an important piece of information concerning the evolution of the replication machine of these large DNA viruses.

Comparisons of putative helicase genes also reveal a

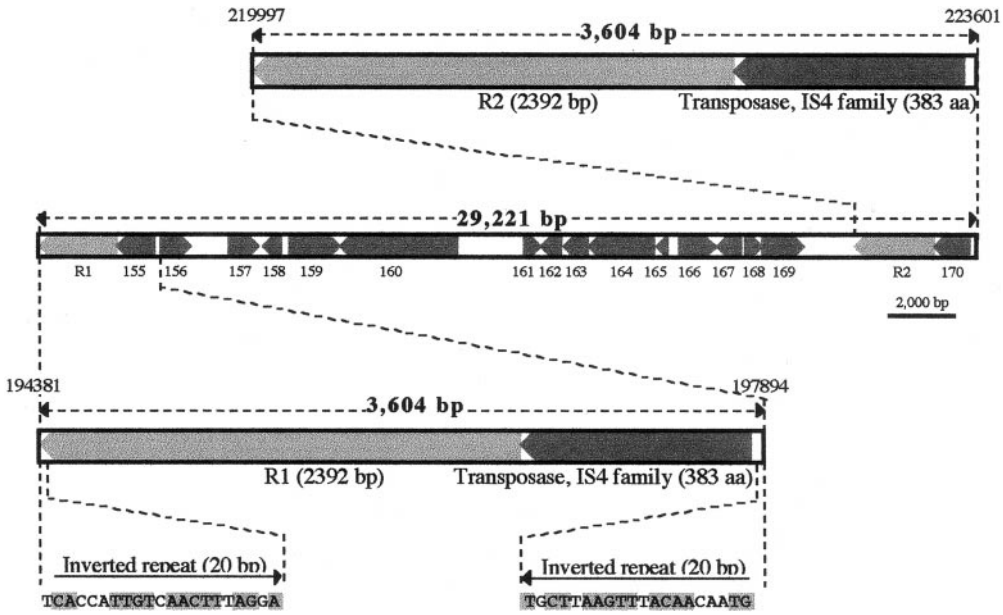

FIG. 7. Organization of a possible viral transposon. (A) Middle: a 29-kb segment bracketed by the repeat elements R1 and R2 linked to putative transposase genes. Top and bottom: detailed structure of R1 and R2 with the sequences of the terminal inverted repeats (shown for element R1 only). A related element R3 is located at nucleotide position 18,807–20,406 of the EsV-1 map and not linked to a transposase gene (see Fig. 5).

curious mixture of eukaryotic and prokaryotic elements. Two EsV-1 genes (ORF 23 and ORF 66) show sequence elements known from eukaryotic DNA/RNA helicases, including the signature DEAD-box motif. In contrast, the putative helicase, encoded by ORF 29, is more related to the RecD subunit of a bacterial RecBCD enzyme that functions as helicase and nuclease in recombination (Myers and Stahl, 1994).

Characteristic postreplicative modifications of *Chlorella* virus genomes are extensive methylations of adenine and cytosine bases (Van Etten and Meints, 1999). This is reflected by the presence of three genes for cytosine DNA methylases and two genes for adenine DNA methylases in the PBCV-1 genome (Van Etten and Meints, 1999). In contrast, the EsV-1 genome appears to lack genes for DNA methylating enzymes even though EsV-1 DNA contains a few methylated bases (Lanka *et al.,* 1993). Interestingly, however, EsV-1 encodes an enzyme with similarities to bacterial 5-methylcytosine-specific restriction nucleases (ORF 16). We speculate that this enzyme together with other putative viral nucleases (ORF 119, ORF 139, ORF 168) may be required for the breakdown of host cell nucleic acids that occurs upon induction of EsV-1 replication in sporangia and gametangia of the infected host (Müller *et al.,* 1998; Wolf *et al.,* 1998, 2000).

Integration. A characteristic feature of the phaeovirus life cycle is the long latency period after infection when the viral genome is mitotically transmitted from cell to cell during development of the host. As suggested by work with EsV-1 (Delaroque *et al.,* 1999) and the related *Feldmannia* sp. Virus (FsV, Van Etten and Meints, 1999), a basis for the stable transmission is most likely the integration of the viral DNA into the host cell genome. In support of this, the EsV-1 genome contains a gene (ORF 213) encoding a large protein with a 200 amino acid region at the carboxyl terminus that shows significant similarities to the catalytic domain of the integrase (Int) family of site-specific recombinases. Integrases cooperate with other proteins to integrate and excise large DNA into and out of the host genome (Landy, 1993). The putative EsV-1 integrase shares with bacteriophage-type integrases the two characteristic homology boxes that Nunes-Düby *et al.* (1998) have identified in their comparisons of more than 100 site-specific recombinases.

The latent state of the EsV-1 genome could be determined by proteins with similarities to bacteriophage regulators of lysogeny as encoded by ORF 117 and ORF 197 (Table 1).

A possible viral transposon. The EsV-1 genome has three dispersed large repeats, termed R1, R2, and R3 (Fig. 7), containing ORFs which we presently do not count as genes because they lack the upstream AT-rich sequence element and do not match any one of the genes in data banks. More interestingly, repeats R1 and R2 are located downstream of putative genes with the potential to code for a transposase of the bacterial IS4 family (IS, insertion sequence; Rezsöhazy *et al.,* 1993). Indeed, the R1- and R2-transposase units are bracketed by imperfect inverted repeats of 20 bp which are also related to the ends of IS4 elements found in bacterial genomes (Fig. 7). Furthermore, R1 and R2 frame a genomic segment of ca. 29 kb and could together constitute one large viral transposon. Similar to bacterial

transposons, the EsV-1 transposon could transmit functions that are advantageous to the host. For example, the putative EsV-1 transposon contains a gene (ORF 169) with the potential to code for a pathogenesis-related 5 protein (PR5), also called thaumatin-like protein, a factor known to be involved in plant defense mechanisms (Hu and Reddy, 1997 and references therein).

Of course, direct experiments are needed to determine whether and how the IS4-related transposases function in EsV-1-infected cells. If they help to boost the defense of the host cells, as the presence of a thaumatin gene suggests, it may be another piece of evidence explaining the close, perhaps symbiotic, relationship between EsV-1 and its host (Müller *et al.,* 1998).

Polysaccharide metabolism. The first EsV-1 gene identified, gp1, was shown to encode a glycoprotein (Klein *et al.,* 1995) and to be highly conserved among phaeoviruses (Müller *et al.,* 1996; Sengco *et al.,* 1996; Maier *et al.,* 1998). Interestingly, recent BLAST searches revealed that the gp1 sequence (ORF 226) has significant homologies with bacterial alginate mannuronan C-5-epimerases (Table 1). This enzyme could modify the structure of the host alginate or could have alginate lyase activity degrading the algal alginate.

In addition to the gp1 gene (ORF 226), at least two other genes may be involved in polysaccharide metabolism. One gene (ORF 83) could encode a GDP-mannose/UDP-glucose dehydrogenase. Members of this dehydrogenase family are known to provide precursors for many glycosyltransferase such as hyaluronan synthase, alginate synthase, and chitin synthase. In fact, the EsV-1 gene with a possible role in polysaccharide metabolism (ORF 84) could code for a protein with similarity to glycosyltransferases (Table 1). Although previously described as a chitin synthase (Müller *et al.,* 1998), it is more likely that the ORF 84 protein, together with the ORF 83 protein, may be somehow involved in alginate synthesis. If so, the gp1 epimerase could modify the alginate product. Surprisingly, the EsV-1 GDP-mannose/UDP-glucose dehydrogenase is similar to bacterial enzymes, whereas the glycosyltransferase is related to eukaryotic enzymes. A fourth gene which most probably belongs to this group of functionally related genes is ORF 166, encoding a protein with a cellulose binding motif.

Protein–protein interactions. The EsV-1 genome encodes several proteins with regions rich in hydrophobic amino acids which together resemble helical transmembrane domains (TM, Table 1). Among these proteins are three of the hybrid kinases (Fig. 6A), and it would be of interest to determine whether these and the other TM-containing proteins are constituents of the viral particle and embedded in a viral lipid layer (Wolf *et al.,* 1998, 2000).

A surprisingly large number of predicted EsV-1 proteins contain classic protein–protein interaction domains such as ankyrin repeats and ring finger domains (some which have been described before for EsV-1 by Delaroque *et al.,* 2000a; and for the related *Feldmannia* sp. Virus by Krueger *et al.,* 1996). In addition, two EsV-1 polypeptides (ORF 80 and ORF 173) have regions that resemble parts of the bacterial DnaJ chaperone protein (Kelley, 1998), and two genes (ORF 28 and ORF 158) encode SET interaction domain originally identified in the *Drosophila* genes Suvr3–9, Enhancer-of-zeste, and Trithorax (Stassen *et al.,* 1995; Tschiersch *et al.,* 1994; Jones and Gelbart, 1993).

It seems that the wide distribution of interaction domains among the viral proteins reflects a complex network between the viral and the host proteins. This network is apparently needed to coordinate the events during the latent and the lytic part of the viral life cycle.

CONCLUSIONS

The EsV-1 sequence invites comparisons with that of the other sequenced phycodnavirus genome, the *Chlorella* virus PBCV-1 (Van Etten and Meints, 1999). The EsV-1 genome and the PBCV-1 genome are quite similar in size (around 330 kb), but differ in organization as EsV-1 DNA contains several tandem and dispersed repeats, whereas the PBCV-1 DNA is mainly composed of single-copy elements. This is one reason EsV-1 DNA has only 231 major protein-encoding genes compared to 376 in the PBCV-1 genome. Only 28 genes in the EsV-1 and the PBCV-1 genome share sequence similarities. These include the genes for the major capsid protein (EsV-1 ORF 116), as well as replication proteins such as DNA polymerase, the processivity factor PCNA, the ATPase, and two helicases.

However, most genes shared by the two viruses have unknown functions. In fact, the genetic differences between the two viral genomes may be more significant than their similarities. Examples are the different sets of enzymes involved in DNA synthesis and DNA modification or in polysaccharide synthesis. Furthermore, EsV-1 appears to rely mainly on hybrid histidine kinases for phosphate transfer reactions, while PBCV-1 has more genes with the potential to code for tyr- and ser-/thr-specific protein kinases. Furthermore, EsV-1 encodes just two proteins for transcriptional regulation, small polypeptides with similarities to $\alpha/\beta/\alpha$ domain of TFIID-18 subunit (ORF 193 and ORF 196), while PBCV-1 has several genes which could be involved in transcriptional and translational regulation. However, the transcription during the EsV-1 life cycle could be regulated through chromatin-remodeling systems such as those containing the polypeptides with SET, POZ/BTB (Broad Complex, Tramtrack, and Bric-a-brac/poxvirus and zinc finger; ORF 40), and BAF60b (ORF 129) domains, which are absent in PBCV-1 (Bardwell and Treisman, 1994; Cairns *et al.,* 1996). Finally, EsV-1 has an elaborate set of

genes for DNA integration and transposition, whereas PBCV-1 just encodes a transposase (Van Etten and Meints, 1999).

Several of these integration/transposition genes of EsV-1 have similarities to corresponding bacteriophage proteins. In fact, the EsV-1 genome can be considered to be an ensemble of genes of bacterial, bacteriophage, and eukaryotic or archaean origins. As just mentioned, the functions for DNA integration are related to bacteriophage enzymes, while functions for polysaccharide synthesis and for DNA restriction are similar to bacterial enzymes, and the hybrid histidine kinases and DNA replication enzymes could be of eukaryotic or archaean origin. How did the EsV-1 genome evolve? Could it be the result of fusions of genomic segments from organisms of three major branches of life? An intriguing alternative possibility is that the EsV-1 DNA as well as the PBCV-1 DNA are remnants of a cellular genome that once lived in symbiosis with their algal hosts, but gradually lost old and gained new genes in response to the different selection pressures exerted on them in their particular environments. This scenario would explain among other features why EsV-1 and PBCV-1 encode different, but overlapping, functions essential for genome replication.

MATERIALS AND METHODS

Genomic sequencing of EsV-1

High molecular EsV-1 DNA was prepared from virus particles embedded in agarose as previously described (Lanka *et al.,* 1993; Klein *et al.,* 1995). DNA (10 μg) was randomly sheared with a nebulizer following the procedure of Pohl and Maier (1995). The fragments were separated on an agarose gel and the selected 1.2-kb fraction isolated using the Qiaex II kit (Qiagen). Sticky ends were filled out using Klenow polymerase, cloned in M13 vectors, and used directly for transformation of competent *E. coli* cells using standard protocols (Sambrook *et al.,* 1989). The plated bacteria were picked into 96-well microtiter plates and stored at room temperature. Plasmid DNA was isolated from the shotgun clones using the Biorobot 9600 (Qiagen) and sequenced manually in 96-well microtiter plates according to the dideoxy-chain termination method of Sanger *et al*. (1977) with the Big Dye Kit (Perkin–Elmer). The sequences were run on an ABI377 HT (Perkin–Elmer) for 4 or 12 h to achieve 600 or 800 bp read length, respectively. The data were extracted using a Sequencing Analysis Software Package (Perkin–Elmer).

Assembly and finishing

The shotgun sequences were assembled with the SeqManII Lasergene Software (DNASTAR Inc.). Raw data were produced until an average coverage of 6.4× in a total of 26 contigs (>1 kb) was reached. The finishing process was carried out in several phases: closing of physical gaps (by performing PCR on genomic DNA with primer pairs positioned on independent contig ends), closing of sequence gaps (by sequencing the complementary strand at a respective site), and manual editing of the entire sequence (with resequencing of particular clones in positions of sequence ambiguities). A total number of 141 oligonucleotides were required for PCR and sequencing. The finishing phase resulted in one final linear contig.

Analysis of sequence data

The open reading frames were identified with the lasergene biocomputing software package (DYNASTAR Inc.). Homology searches were carried out with the BLAST program (Altschul *et al.,* 1990, scoring matrix blosum 62) against the nonredundant protein databases at the NCBI. Protein motifs were searched against the SMART (Schultz *et al.,* 2000), PROSITE (Hofmann *et al.,* 1999), and Pfam (Bateman *et al.,* 2000) databases. Charged amino acid clusters and repeats were found by using the SAPS program of Brendel *et al.* (1992). The presence of putative signal peptides (Nielsen *et al.,* 1997), transmembrane domains (Fasman and Gilbert, 1990; von Heijne, 1992), and coiled coil domains (Lupas *et al.,* 1991) were identified. The intergenic regions were also searched against DNA databases using the BLAST computer program (Altschul *et al.,* 1990).

Nucleotide sequence accession number

The EsV-1 genome sequence has been deposited in GenBank under the Accession No. AF204951.

ACKNOWLEDGMENTS

We thank Ingo Maier for critical reading of the manuscript and an anonymous reviewer for suggestions on the EsV-1 DNA ends.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410.

Bardwell, V. J., and Treisman., R. (1994). The POZ domain: A conserved protein-protein interaction motif. *Genes Dev.* **15,** 1664–1677.

Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Howe, K. L., and Sonnhammer, E. L. L. (2000). The Pfam protein families database. *Nucleic Acids Res.* **28,** 263–266.

Bawden, A. L., Glassberg, K. J, Diggans, J., Shaw, R., Farmerie, W., and Moyer, R. W. (2000). Complete genomic sequence of the *Amsacta moorei* entomopoxvirus: Analysis and comparison with other poxviruses. *Virology* **274,** 120–139.

Bräutigam, M., Klein, M., Knippers, R., and Müller, D. G. (1995). Inheritance and meiotic elimination of a virus genome in the host *Ectocarpus siliculosus* (Phaeophyceae). *J. Phycol.* **31,** 823–827.

Brendel, V., Bucher, P., Nourbakhsh, I., Blaisdell, B. E., and Karlin, S. (1992). Methods and algorithms for statistical analysis of protein sequences. *Proc. Natl. Acad. Sci. USA* **89,** 2002–2006.

Cairns, B. R., Levinson, R. S., Yamamoto, K. R., and Kornberg, R. D.

(1996). Essential role of Swp73p in the function of yeast Swi/Snf complex. *Genes Dev.* **10,** 2131–2144.

Davis, S. J., Vener, A. V., and Vierstra, R. D. (1999). Bacteriophytochromes: Phytochrome-like photoreceptors from nonphotosynthetic eubacteria. *Science* **286,** 2517–2520.

Delaroque, N., Maier, I., Knippers, R., and Müller, D. G. (1999). Persistent virus integration into the genome of its algal host, *Ectocarpus siliculosus* (Phaeophyceae). *J. Gen. Virol.* **80,** 1367–1370.

Delaroque, N., Wolf, S., Müller, D. G., and Knippers, R. (2000a). Characterization and immunolocalization of major structural proteins in the brown algal virus EsV-1. *Virology* **269,** 148–155.

Delaroque, N., Wolf, S., Müller, D. G., and Knippers, R. (2000b). The brown algal virus EsV-1 particle contains a putative hybrid histidine kinase. *Virology* **273,** 383–390.

Fasman, G. D., and Gilbert, W. A. (1990). The prediction of transmembrane protein sequences and their conformation: An evaluation. *Trends Biochem. Sci.* **15,** 89–92.

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

Grebe, T. W., and Stock, J. B. (1999). The histidine protein kinase superfamily. *Adv. Microb. Physiol.* **41,** 139–227.

Henry, E. C., and Meints, R. H. (1994). Recombinant viruses as transformation vectors of marine macroalgae. *J. Appl. Phycol.* **6,** 247–253.

Hoch, J. A., and Silhavy, T. J. (1995). "Two-Component Signal Transduction." ASM Press, Washington, D.C.

Hofmann, K., Bucher, P., Falquet, L., and Bairoch, A. (1999). The PROSITE database, its status in 1999. *Nucleic Acids Res.* **27,** 215–219.

Hübscher, U., Nasheuer, H. P., and Syvaoja, J. E. (2000). Eukaryotic DNA polymerases, a growing family. *Trends Biochem. Sci.* **25,** 143–147.

Hu, X., and Reddy, A. S. N. (1997). Cloning and expression of a PR5-like protein from *Arabidopsis*: Inhibition of fungal growth by bacterially expressed protein. *Plant Mol. Biol.* **34,** 949–959.

Jones, R. S., and Gelbart, W. M. (1993). The *Drosophila* polycomb-group gene enhancer of zeste contains a region with sequence similarity to trithorax. *Mol. Cell. Biol.* **13,** 6357–6366.

Kelley, W. L. (1998). The J-domain family and the recruitment of chaperone power. *Trends Biochem. Sci.* **23,** 222–227.

Klein, M., Lanka, S., Müller, D. G., and Knippers, R. (1994). Single-stranded regions in the genome of *Ectocarpus siliculosus*. *Virology* **202,** 1076–1078.

Klein, M., Lanka, S. T. J., Knippers, R., and Müller, D. G. (1995). Coat protein of the *Ectocarpus siliculosus* virus. *Virology* **206,** 520–526.

Kotani, H., and Tabata, S. (1998). Lessons from sequencing of the genome of a unicellular cyanobacterium, *Synechocystis* sp. PCC6803. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **49,** 151–171.

Krueger, S. K., Irvey, R. G., Henry, E., and Meints, R. H. (1996). A brown algal virus genome contains a "RING" zinc finger motif. *Virology* **219,** 301–303.

Kutish, G. F., Li, Y., Lu, Z., Furuta, M., Rock, D. L., and Van Etten, J. L. (1996). Analysis of 76 kb of the *Chlorella* virus PBCV-1 330 kb genome map positions 182 to 258. *Virology* **223,** 303–317.

Kuzio, J., Pearson, M. N., Harwood, S. H., Funk, C. J., Evans, J. T., Slavicek, J. M., and Rohrmann, G. F. (1999). Sequence and analysis of the genome of a baculovirus pathogenic for *Lymantria dispar*. *Virology* **253,** 17–34.

Landy, A. (1993). Mechanistic and structural complexity in the site-specific recombination pathways of Int and FLP. *Curr. Opin. Genet. Dev.* **3,** 699–707.

Lanka, S. T. J., Klein, M., Ramsperger, U., Müller, D. G., and Knippers, R. (1993). Genome structure of a virus infecting the marine brown alga *Ectocarpus siliculosus*. *Virology* **193,** 802–811.

Lavrukhin, O. V., Fortune, J. M., Wood, T. G., Burbank, D. E., Van Etten, J. L., Osheroff, N., and Lloyd, R. S. (2000). Topoisomerase II from *Chlorella* virus PBCV-1. Characterization of the smallest known type II topoisomerase. *J. Biol. Chem.* **275,** 6915–6921.

Lee, A. M., Ivey, R. G., and Meints, R. H. (1998a). Repetitive DNA insertion in a protein kinase ORF of a latent FsV (*Feldmannia* sp. Virus) genome. *Virology* **248,** 35–45.

Lee, A. M., Ivey, R. G., and Meints, R. H. (1998b). The DNA polymerase gene of a brown algal virus: Structure and phylogeny. *J. Phycol.* **34,** 608–615.

Lupas, A., Van Dyke, M., and Stock, J. (1991). Predicting coiled coils from protein sequences. *Science* **252,** 1162–1164.

Maier, I., Wolf, S., Delaroque, N., Müller, D. G., and Kawai, H. (1998). A DNA virus infecting the marine brown alga *Pilayella littoralis* (Ectocarpales, Phaeophyceae) in culture. *Eur. J. Phycol.* **33,** 213–220.

Miyata, S., Urao, T., Yamaguchi-Shinozaki, K., and Shinozaki, K. (1998). Characterization of genes for two-component phosphorelay mediators with a single HPt domain in *Arabidopsis thaliana*. *FEBS Lett.* **437,** 11–14.

Mizuno, T. (1997). Compilation of all genes encoding two-component phosphotransfer signal transducers in the genome of *Escherichia coli*. *DNA Res.* **4,** 161–168.

Mossi, R., and Hübscher, U. (1998). Clamping down on clamps and clamp loaders-the eukaryotic replication factor C. *Eur. J. Biochem.* **254,** 209–216.

Müller, D. G. (1991a). Mendelian segregation of a virus genome during host meiosis in the marine brown alga *Ectocarpus siliculosus*. *J. Plant Physiol.* **137,** 739–743.

Müller, D. G. (1991b). Marine virioplankton produced by infected *Ectocarpus siliculosus* (Phaeophyceae). *Mar. Ecol. Prog. Ser.* **76,** 101–102.

Müller, D. G., Bräutigam, M., and Knippers, R. (1996). Virus infection and persistence of foreign DNA in the marine brown alga *Feldmannia simplex* (Ectocarpales, Phaeophyceae). *Phycologia* **35,** 61–63.

Müller, D. G., Kapp, M., and Knippers, R. (1998). Viruses in marine brown algae. *Adv. Virus Res.* **50,** 49–67.

Müller, D. G., Westermeier, R., Morales, J., Garcia Reina, G., del Campo, E., Correa, J. A., and Rometsch, E. (2000). Massive prevalence of viral DNA in *Ectocarpus* (Phaeophyceae, Ectocarpales) from two habitats in the North Atlantic and South Pacific. *Bot. Mar.* **43,** 157–159.

Myers, R. S., and Stahl, F. W. (1994). χ and the RecBCD enzyme of *Escherichia coli*. *Annu. Rev. Genet.* **28,** 49–70.

Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10,** 1–6.

Nunes-Düby, S. E., Kwon, H. J., Tirumalai, R. S., Ellenberger, T., and Landy, A. (1998). Similarities and differences among 105 members of the Int family of site-specific recombinases. *Nucleic Acids Res.* **26,** 391–406.

Plugge, B., Gazzarini, S., Nelson, M., Cerana, R., Van Etten, J. L., Derst, C., DiFrancesco, D., Moroni, A., and Thiel, G. (2000). A potassium channel protein encoded by *Chlorella* virus PBCV-1. *Science* **287,** 1641–1644.

Pohl, T. M., and Maier, E. (1995). Sequencing 500 kb of yeast DNA using a GATC 1500 direct blotting electrophoresis system. *Biotechniques* **19,** 482–486.

Pringle, C. R. (1998). Virus taxonomy, San Diego 1998. *Arch. Virol.* **143,** 1449–1459.

Rezsöhazy, R., Hallet, B., Delcour, J., and Mahillon, J. (1993). The IS4 family of insertion sequences: Evidence for a conserved transposase motif. *Mol. Microbiol.* **9,** 1283–95.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). "Molecular Cloning: A Laboratory Manual," 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sanger, F., Nicklen, S., and Coulsen, A. R. (1977). DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.

Schultz, J., Copley, R. R., Doerks, T., Ponting, C. P., and Bork, P. (2000). SMART: A Web-based tool for the study of genetically mobile domains. *Nucleic Acids Res.* **28,** 231–234.

Schuster, A. M., Graves, M., Korth, K., Ziegelbein, M., Brumbaugh, J., Grone, D., and Meints, R. H. (1990). Transcription studies of a 4.3-kbp

fragment from a ds-DNA eukaryotic algal virus. *Virology* **176,** 515–523.

Sengco, M. R., Bräutigam, M., Kapp, M., and Müller, D. G. (1996). Detection of virus DNA in *Ectocarpus siliculosus* and *E. fasciculatus* (Phaeophyceae) from various geographic areas. *Eur. J. Phycol.* **31,** 73–78.

Stassen, M. J., Bailey, D., Nelson, S., Chinwalla, V., and Harte, P. J. (1995). The *Drosophila* trithorax proteins contain a novel variant of the nuclear receptor type DNA binding domain and an ancient conserved motif found in other chromosomal proteins. *Mech. Dev.* **52,** 209–223.

Stock, A. M., Robinson, V. L., and Goudreau, P. N. (2000). Two-component signal transduction. *Annu. Rev. Biochem.* **69,** 183–215.

Suzuki, T., Imamura, A., Ueguchi, C., and Mizuno, T. (1998). Histidine-containing phosphotransfer (HPt) signal transducers implicated in His-to-Asp phosphorelay in *Arabidopsis. Plant Cell Physiol.* **12,** 1258–1268.

Tschiersch, B., Hofmann, A., Krauss, V., Dorn, R., Korge, G., and Reuter, G. (1994). The protein encoded by the *Drosophila* position-effect variegation suppressor gene Su(var)3–9 combines domains of antagonistic regulators of homeotic gene complexes. *EMBO J.* **13,** 3822–3831.

Van Etten, J. L., and Meints, R. H. (1999). Giant viruses infecting algae. *Annu. Rev. Microbiol.* **53,** 447–494.

Van Etten, J. L., Lane, L. C., and Meints, R. H. (1991). Viruses and virus-like particles of eukaryotic algae. *Microbiol. Rev.* **55,** 586–620.

Villarreal, L. P., and DeFilippis, V. R. (2000). A hypothesis for DNA viruses as the origin of eukaryotic replication proteins. *J. Virol.* **74,** 7079–7084.

Von Heijne, G. (1992). Membrane protein structure prediction, hydrophobicity analysis and the positive-inside rule. *J. Mol. Biol.* **225,** 487–494.

Wolf, S., Maier, I., Katsaros, C., and Müller, D. G. (1998). Virus assembly in *Hincksia hincksiae* (Ectocarpales, Phaeophyceae). An electron and fluorescence microscopic study. *Protoplasma* **203,** 153–167.

Wolf, S., Müller, D. G., and Maier, I. (2000). Assembly of a large icosahedral DNA virus, MclaV-1, in the marine alga *Myriotrichia clavaeformis* (Dictyosiphonales, Phaeophyceae). *Eur. J. Phycol.* **35,** 163–171.

Yáñez, R. J., Rodríguez, J. M., Nogal, M. L., Yuste, L., Enríquez, C., Rodriguez, J. F., and Viñuela, E. (1995). Analysis of the complete nucleotide sequence of African swine fever virus. *Virology* **208,** 249–278.

Yeh, K. C., Wu, S. H., Murphy, J. T., and Lagarias, J. C. (1997). A cyanobacterial phytochrome two-component light sensory system. *Science* **277,** 1505–1508.