

# A novel evolutionary strategy revealed in the Phaeoviruses

Article

**Published Version** 

Creative Commons: Attribution 4.0 (CC-BY)

**Open Access** 

Stevens, K., Weynberg, K., Bellas, C., Brown, S., Brownlee, C., Brown, M. T. and Schroeder, D. C. (2014) A novel evolutionary strategy revealed in the Phaeoviruses. PLoS ONE, 9 (1). e86040. ISSN 1932-6203 doi: https://doi.org/10.1371/journal.pone.0086040 Available at http://centaur.reading.ac.uk/79417/

It is advisable to refer to the publisher's version if you intend to cite from the work.

Published version at: http://dx.doi.org/10.1371/journal.pone.0086040 To link to this article DOI: http://dx.doi.org/10.1371/journal.pone.0086040

Publisher: Public Library of Science

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the <u>End User Agreement</u>.

www.reading.ac.uk/centaur

CentAUR



### Central Archive at the University of Reading

Reading's research outputs online

## A Novel Evolutionary Strategy Revealed in the Phaeoviruses

## Kim Stevens<sup>1,2</sup>, Karen Weynberg<sup>1,2</sup>, Christopher Bellas<sup>1,2</sup>, Sonja Brown<sup>1,2</sup>, Colin Brownlee<sup>1,2</sup>, Murray T. Brown<sup>2</sup>, Declan C. Schroeder<sup>1\*</sup>

1 Cell and Molecular Department, Marine Biological Association, Plymouth, Devon, United Kingdom, 2 School of Marine Science and Engineering, University of Plymouth, Plymouth, Devon, United Kingdom

#### Abstract

Phaeoviruses infect the brown algae, which are major contributors to primary production of coastal waters and estuaries. They exploit a Persistent evolutionary strategy akin to a *K*- selected life strategy via genome integration and are the only known representatives to do so within the giant algal viruses that are typified by *r*- selected Acute lytic viruses. In screening the genomes of five species within the filamentous brown algal lineage, here we show an unprecedented diversity of viral gene sequence variants especially amongst the smaller phaeoviral genomes. Moreover, one variant shares features from both the two major sub-groups within the phaeoviruses. These phaeoviruses have exploited the reduction of their giant dsDNA genomes and accompanying loss of DNA proofreading capability, typical of an Acute life strategist, but uniquely retain a Persistent life strategy.

Citation: Stevens K, Weynberg K, Bellas C, Brown S, Brownlee C, et al. (2014) A Novel Evolutionary Strategy Revealed in the Phaeoviruses. PLoS ONE 9(1): e86040. doi:10.1371/journal.pone.0086040

Editor: Stephen J. Martin, Sheffield University, United States of America

Received September 16, 2013; Accepted December 4, 2013; Published January 21, 2014

**Copyright:** © 2014 Stevens et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This research was funded by INTERREG programme France (Channel) – England, MARINEXUS (Ref: 1956/4073) and Plymouth University. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

\* E-mail: dsch@mba.ac.uk

#### Introduction

All viruses broadly follow one of two life strategies, Acute or Persistent [1,2,3]. Moreover, the switch from Persistent to Acute in animal systems underpins emerging new viral epidemiology, notable examples being influenza, measles and HIV [2]. This transformation is often triggered by viruses jumping from one species to another. Viruses that follow an Acute life strategy have characteristic features that associate them with a disease phenotype; high reproduction and mutation rates, and greater dependency on host population densities for transmission. Many animal viral infections that are responsible for emerging epidemic diseases follow this Acute infection dynamic that originated from a Persistent viral life strategist [2]. Despite their likely prevalence, Persistent viral life strategies are not well described. Persistence is defined as a stable coexistence in an individual host, seldom causing disease, and transmission is often from parent to offspring [1]. Phaeoviruses infect the Ectocarpales brown algae, which are major contributors to primary production of coastal waters and estuaries [4], and separated from the kelps around 100 Ma [5]. Viral infections in protists contribute significantly to the sheer abundance of viruses in our oceans [6], and have been shown to play important roles in some of the major oceanic processes, such as plankton mortality [7,8], nutrient cycling and carbon storage [9,10]. Their ubiquitous nature means that viruses affect every aspect of life in the marine environment, and their importance in such fundamental areas as evolution [3,11], the global food web and even climate change should not be underestimated [3,9,12].

Protist viruses belonging to the family *Phycodnaviridae* [13] are members of the wider grouping of nuclear cytoplasmic large

dsDNA viruses (NCLDVs). The coccolitho-[11,14] and phaeoviruses [4,15] are two examples of NCLDVs having opposing life strategies Acute vs Persistent, respectively. The former are lytic algal bloom terminators [14], while the latter covertly infect and integrate their genomes via the gamete and/or spore life stages of the host, forming a latent provirus which is transmitted to all cells during adult development [4]. As with most persistent viruses, phaeoviruses have no noticeable negative impact on the life-cycle of the host; however, overt symptoms of phaeovirus infection can be seen when the reproductive organs become deformed and produce virions, instead of gametes or spores (Figure 1).

To date, phaeovirus identity has only been confirmed for viruses infecting three species of filamentous brown algae: Ectocarpus siliculosus (Dillwyn) Lyngbye (Esil), Feldmannia sp. and Feldmannia irregularis (Kützing) Hamel (Firr); infected by EsV-1, FsV and FirrV-1, respectively [16]. They vary in genome size from 180-336 kb (Table 1). In addition, the genome of an Ectocarpus strain was found to contain a transcriptionally inactive copy of an EsV-1like provirus [4]. Complete genome sequences show that EsV-1, FirrV-1 and FsV-158 contain a limited number of common single copy core genes, as well as many unique genes [15]. Five phaeoviruses, identified by morphology and life cycle, infecting Ectocarpus fasciculatus (Harvey) (Efas), Feldmannia simplex (Crouan & Crouan) Hamel (Flex), Hincksia hincksiae (Harvey) Silva (Hinc), Pylaiella littoralis (Linnaeus) Kjellman (Plit) and Myriotrichia clavaeformis (Harvey) (Mcla) have also been described in the literature (Table 1) [16]. Here we report on the phylogenetic placement of these phaeoviruses, using single and multi-gene phylogenies for three NCLDV core single copy genes, namely the major capsid protein (MCP), DNA polymerase (DNApol) genes,



**Figure 1. Epifluorescence microscope images of** *E. siliculosus.* The pink stained individual spores (combination of DAPI stained blue DNA and red auto-fluorescence from nuclei and chloroplasts, respectively) are clearly visible within the normal zoidangium (A), whereas in (B) the zoidangium is misshapen and heavily stained showing that the space is filled with densely packed blue viral particles. doi:10.1371/journal.pone.0086040.g001

and a hitherto untested viral superfamily III helicase (VACV D5-like) gene.

#### **Results and Discussion**

EsV-1 is the only virus known to infect *E. siliculosus*, while Ivey *et al.*[17] reported the presence of two (and potentially four) different size variants (158 kb and 178 kb) of phaeoviruses in cultures of *Feldmannia* sp. Delaroque *et al.* [18] reported an incomplete FirrV-1 genome, with no evidence of multiple variants, within *F. irregularis*. Our viral sequences for EsV-1 (Table 1 & Figure 2). Notably, no additional sequence variation for EsV-1 could be found. The other available DNApol gene sequences for the reference genome, FirrV-1, were identified within the *Feldmannia irregularis* (Firr 1) isolate (Figure 2); however, at least one additional variant could also be identified (Table 1). This result is the likely explanation for the inability of Delaroque *et al.* to assemble the FirrV-1 genome [18]. All the other ectocarpoid

strains contained two or more viral sequence variants, with the *Feldmannia simplex* (Flex) isolate containing at least eight different variants (Table 1). Our Bayesian and Maximum Likelihood inference trees (DNApol or multigene) were largely in agreement that the phaeovirus sequence variants group should be split into two distinct sub-groups: a virus sub-group A that infect multiple species across three families of the Ectocarpales (Figures 3 & 4) and a second sub-group B containing members that infect the genus *Feldmannia*.

Furthermore, there are two unexpected observations from these phylogenies. Firstly, the Flex 8 variant shares features with both of the sub-groups, whilst, unsurprisingly, being more closely connected to sub-group B (Figure 3 & Figure 4). A closer look at the DNApol sequence (Figure 2) shows not only the overall conservation of amino acids (32%) across all the phaeoviruses and the wider eukaryote kingdom as a whole, but also how certain amino acids can be assigned to either sub-group A (triangles, Figure 2) or sub-group B (inverted triangles, Figure 2). Moreover, one important conserved region, Pol III dNTP binding site, is

Table 1. Ectocarpoid strains used for phaeovirus screening (adapted from Schroeder [16]).

Strain	Species	Family	Location	Genome	Number of	sequence v	Concatenations**	
				kb	DNApol	МСР	Helicase	
Esil	Ectocarpus siliculosus	Ectocarpaceae	New Zealand	336	1 (1)	1 (1)	1 (1)	1
Efas	Ectocarpus fasciculatus	Ectocarpaceae	France	320	2 (2)	1 (1)	2 (2)	2
Plit	Pylaiella littoralis	Acinetosporaceae	Alaska	280	1 (1)	1 (1)	1 (1)	1
Hinc	Hincksia hincksiae	Acinetosporaceae	France	240	1 (1)	-	2 (1)	-
Mcla	Myriotrichia clavaeformis	Chordariaceae	Argentina	320	1 (1)	2 (2)	-	2
Firr	Feldmannia irregularis	Acinetosporaceae	Canary Islands	180	2 (2)	3 (2)	2 (2)	4
Flex	Feldmannia simplex.	Acinetosporaceae	Ireland	220	9 (8)	6 (4)	8 (3)	22

\*: variant in DNA sequence (HG003317 - HG003355) with amino acid variation indicated in parentheses. A negative PCR result is indicated by a minus symbol. \*\*: possible permutations for DNApol and MCP as seen in Figure 4.

doi:10.1371/journal.pone.0086040.t001

#### Pol II (dNTP binding site)

	470	480	490	500	510	520	530	540	550	560	570
		]	[]]		.	]					
EsV-1 genome	MSLILNRIHGEYVCN	YAAAKKKMAA	DGKQVLNEG	KGASVIDAKK	GFYEKDPIVTN	<b>DFASLYPSI</b>	MRLKQLCYTTI	V-R	DV	-KYRGIEGVNY	EDHQ
Esil 1	MSLILNRIHGEYVCN	YAAAKKKMAA	DGKQVLNEG	KGASVIDAKK	GFYEKDPIVTN	DFASLYPSI	MRLKQLCYTTI	V-R	DV	-KYRGIEGVNY	EDHQ
Efas 1	MSLILNRIHGEYVCN	YAAAKKKMAA	DGTQVLNDG	KGASVIDAKK	GFYEQDPVVTM	<b>DFASLYPSI</b>	MRLKQLCYTTI	V-K	DD	-KYRGIEGIVY	EDHQ
Efas 2	MSLILNRIHGEYVCN	YAVAKKKVA	DGKQVLNEG	KGASVIDAKK	GFYEKDPVVTN	DFASLYPSI	MRLKQLCYTTI	V-K	DV	-KYRDIKGVTY	EDHQ
Esil genome provirus	MSLILNRIHGEYVCN	YVVAKKKVA	DGKQVLNEG	KGASVIDAKK	GFYEKDPVVTN	DFASLYPSI	MRLKQLCYTTI	V-K	<b>D</b> V	-KYRGIEGVTY	EDHQ
Hinc 1	MSLILNRIHGEFVCN	YSPVKKKVA	DGNEVLLEG	KGASVIEAKK	GFYEKDPVVTN	DFASLYPSI	MRLKOLCYTTI	V-K	<b>D</b> L	-KYRGIEGIVY	EDHE
Mcla 1	MSLI?NRTHGEFVCN	YMPAKRKTT/	DGKQVPNEG	KGAS?IDAKK	GFYEKDPIVTN	D?NSLYPSI	MRLKQLCYTTI	V-R	<b>D</b> S	-KYRAIEGVSY	EDHE
Plit 1	?SLILNRTHGEFVCN	YTPE-KRVEV	DGKQVPKDG	KGASVIDAKK	GFYETDPVVTN	DFESLYPSI	MRLKQLCYTTE	V-K	ED	-KYRGIDGVVY	EDHE
Flex 8	FSLILDRIFGEYVCN	QLDARAGDG-	GG	TGATVIDAQT	GFHVDDPVVCI	DFASLYPSI	MRWKNLCYTTY	VI	DD	-EFRGIGGVNY	KEFE
FsV-158 genome	LSLILDRVHGEFVCN	KASRVLPGGV	7K	QGATVIDATK	GFHNKDPVVCI	LDFASLYPSI	IRWKNLCYTTY	L	DSD-	-EFANIPGVHY	ERFE
FirrV-1 genome	LSLILDRIHSEYVLN	RSRALEQOD	20CK	QGATVISAKK	GFHCDDPVVCI	LDFASLYPSI	IRWKNLCYTTH	v	DSD-	-EFLDIDGVDY	EKFE
Firr 1	LSLILDRIHSEYVLN	RSRALEQQD	2QCK	QGATVISAKK	GFHCDDPVVCI	LDFASLYPSI	IRWKNLCYTTH	v	DSD-	-EFLDIDGVDY	EKFE
Firr 2	LSLILDRIHSEYVLN	KTRDSTNKQ	DATVSRK	QGATVISAKK	GFHSEDPVVCI	DFASLYPSI	IRWKNLCYTTH	v	NSD-	-EFLGIEGVEY	EKFE
Flex 2	N	YVAKCGDAST	SSKSK	QGATVIDAQK	GFHCEDPVVCI	LDFASLHPSI	IRWKNLCYTTY	IIPNSSSSSS	SSSDSNSDS	GDYLNIPGVMY	EKFE
Flex 7	LSLILDRIHSVYVLN	YVAKCGDAST	CSSKSNSK	QGATVIDAQK	GFHCEDPVVCI	DFASLYPSI	IRWKNLCYTTY	IIPNSS	SNSDS	-DYLNIQGVMY	EKFE
Flex 6	LSLILDRIHSEYVLN	KTAKAEDGSS	SSK	QGATVIDAKK	GFHCEDPVVCI	DFASLYPSI	IRWKNLCYTTY	v	NSD-	-EYLDIPGVVY	EKFE
Flex 5	LSLILDRIHSEYVLN	YTSKTGDTPI	HSK	QGATVIDAQK	GFHCEDPVVCI	LDFASLYPSI	IRWKNLCYTTY	VN	SN	-EYSNIPGVEY	ERFE
Flex 3	LSLILDRIHAEYVLN	TSKTRSDD	CSSHSK	QGATVIDAQK	GFHCEDPVVCI	DFASLYPSI	IRWKNLCYTTY	AN	SN	-EYSSIPGVEY	ERFE
Flex 4	LSLILDRIHSEYVLN	YTSKTRSDD	CSSHSK	QGATVIDAQK	GFHCEDPVVCI	LDFASLYPSI	IRWKNLCYTTY	VN	SN	-EYSSIPGVEY	ERFE
Flex 1	LSLILDRIHADYVLN	YTSKTSND-7	CSSHSK	QGATVIDAQK	GFHCEDPVVCI	DFASSYPSI	IRWKNLCYTTY	VN	SN	-EYSSIPGVEY	ERFE
		Pol III (dNTP binding site)									
		Pol	VI			Pol	III (dNTP bi	nding site)			
	580	<b>Pol</b>	VI 600	610	620	<b>Pol</b>	III (dNTP bin 640	nding site) 650	660	670	
	580	Pol 590	VI 600	610 	620 •••• • <u>••• </u>	<b>Pol</b> 630	III (dNTP bin 640	nding site) 650	660 	670   .	
EsV-1 genome	580    IS <mark>DGVSVTFAHRPG</mark> S	590	000	610    KLAKSEKD-PF	620 	630	III (dNTP bin 640 	nding site) 650 	660 STGRDMI R	670   . TKEYAEKEHGC	 IVVI
EsV-1 genome Esil 1	580    ISDGVSVTFAHRPGS ISDGVSVTFAHRPGS	Pol 590  RSILCELEEN RSILCELEEN	VI 600 	610    TLIMKSEKD-PF.	620 	Pol 630 	III (dNTP bin 640 	nding site) 650   PLVEIAAAVI PLVEIAAAVI	660 STGRDMIKR	670   . TKEYAEKEHGC TKEYAEKEHGC	 NVI NVI
EsV-1 genome Esil 1 Efas 1	580    ISDCVSVTFARPGS ISDCVSVTFARPGS ISDEAIVTFARPGS	Pol 590   RSILCELEEN RSILCELEEN RSILCELEEN	OO 600 L.C.I.I.I.I LGEERKATKH LGEERKATKH	610 LINKSEKD-PF LINKSEKD-PF LINKSEKD-PF	620 Ayslldskoka Ayslldskoka Ayslldskoka	630 630 QKVTMNSIY QKVTMNSIY	III (dNTP bin 640 GFTGTVNNGMI GFTGTVNNGMI GFTGTVNNGMI	nding site) 650 	660    STGRDMIKR STGRDMIKR	670   . TKEYAEKEHGC TKEYAEKEHGC TKEYAEKEHGC	IVN IVN IVN
EsV-1 genome Esil 1 Efas 1 Efas 2	580    ISDGVSVTFAHRPGS ISDGVSVTFAHRPGS ISDGVSVTFAHRPGS	Pol 590 RSILCELEEN RSILCELEEN RSILCELEEN	VI 600 LCEERKATKH LCEERKATKH LCEERKATKH LCEERKATKH	610 LIAKSEKD-PF LIAKSEKD-PF LIAKSEKD-PF LIAKSEKD-PF	620 AYSLLDSKQKA AYSLLDSKQKA AYSLLDSKQKA AYSLLDSKQKA	Fol 630 AQKVTMNSIY AQKVTMNSIY AQKVTMNSIY AQKVTMNSIY	III (dNTP bin 640 GFTGTVNNGMI GFTGTVNNGMI GFTGTVNNGMI GFTGTVNNGMI	nding site) 650 	660 STGRDMIKR STGRDMIKR STGRDMIKR STGRDMIKR	670 TKEYAEKEHGC TKEYAEKEHGC TKEYAEKEHGC TKEYAEKEHGC	IVN IVN IVN IVN
EsV-1 genome Esil 1 Efas 1 Efas 2 Esil genome provirus	580 	Pol 590 RESILCELEEN RESILCELEEN RESILCELEEN RESILCELEEN	VI 600 LEERKATKR LEERKATKR LEERKATKR LEERKATKR LEERKATKR	610    LIMKSEKD-PF LIMKSEKD-PF LIMKSEKD-PF LIMKSEKD-PF	620 AYSLLDSKQKA AYSLLDSKQKA AYSLLDSKQKA AYSLLDSKQKA	Pol 630 AQKVTMNSIY AQKVTMNSIY AQKVTMNSIY AQKVTMNSIY	III (dNTP bin 640 GFTGTVNNGMI GFTGTVNNGMI GFTGTVNNGMI GFTGTVNNGMI	nding site) 650 	660 STGRDMI KR STGRDMI KR STGRDMI KR STGRDMI KR STGRDMI KR	670 	IVN IVN IVN IVN IVN
EsV-1 genome Esil 1 Efas 1 Efas 2 Esil genome provirus Hinc 1	580 	Pol	VI 600 ILGEERKATKR ILGEERKATKR ILGEERKATKR ILGEERKATKR ILGEERKATKR	610    KLAKSEKD-PF LLAKSEKD-PF LLAKSEKD-PF LLAKSEKD-PF LLAKSEKD-PF	620 AYSLLDSKQKA AYSLLDSKQKA AYSLLDSKQKA AYSLLDSKQKA AYSLLDSKQKA	Pol 630  QKVTMNSIY QKVTMNSIY QKVTMNSIY QKVTMNSIY	III (dNTP bin 640 	nding site) 650 	660 STGRDMI R STGRDMI R STGRDMI R STGRDMI R STGRDMI R STGRNMI R	670 TKEYAEKEHGCI TKEYAEKEHGCI TKEYAEKEHGCI TKEYAEKEHGCI TKEYAEKEHGCI TKEYAEKEHGCI	NVI NVI NVI NVI NVI NVI
ESV-1 genome Esil 1 Efas 1 Efas 2 Esil genome provirus Hinc 1 Mcla 1	580    ISDGVSVTFAHRPGS ISDGVSVTFAHRPGS ISDGVSVTFAHRPGS ISDGVSVTFAHRSGS VSEGVATFAHRPGS VSEGVATFAHRPGS	590 SUCCELEEN RSILCELEEN RSILCELEEN RSILCELEEN KSILCELEEN KSILCELEEN	VI 600 ILGEERKATKR ILGEERKATKR ILGEERKATKR ILGEERKATKR ILGEERKATKR ILGEERKATKR	610 	620 AYSLD KQK/ AYSLD KQK/ AYSLD KQK/ AYSLD KQK/ AYSLD KQK/ AYSLD KQK/	Pol 630 AQKVTMNSIY AQKVTMNSIY AQKVTMNSIY AQKVTMNSIY AQKVTMNSIY AQKVTMNSIY	III (dNTP bit 640 GFTGTVNNCMI GFTGTVNNCMI GFTGTVNNCMI GFTGTVNNCMI GFTGTVNNCMI	nding site) 650 PLVEIAAAVI PLVEIAAAVI PLVEIAAAVI PLVEIAAAVI PLVEIAAAVI PLVEIAAAVI	660 STGRDMI KR STGRDMI KR STGRDMI KR STGRDMI KR STGRDMI KR STGRDMI KR	670 TKEYAEKEHGC TKEYAEKEHGC TKEYAEKEHGC TKEYAEKEHGC TKEYAEKEHGC TKEYAEKEHGC TKEYAETEHGC	IVN IVN IVN IVN IVN VVN IVN
EsV-1 genome Esil 1 Efas 1 Efas 2 Esil genome provirus Hinc_1 Mcla 1 Plit 1	580 IIISDGVSVTFAHRPGS ISDGVSVTFAHRPGS ISDGVSVTFAHRPGS ISDGVSVTFAHRPGS VSEGVJATFAHRPGS VSEGVJATFAHRPGS	Pol	VI 600  ILGEERKATKI ILGEERKATKI ILGEERKATKI ILGEERKATKI ILGEERKATKI ILGEERKATKI	610 LLMSEKD-PF LLMSEKD-PF LLMSEKD-PF LLMSEKD-PF LLMSEKD-PF LLMSEKD-PF RMKCETD-PF	620 AYSLLD KQK/ AYSLLD KQK/ AYSLLD KQK/ AYSLLD KQK/ AYSLLD KQK/ AYSLLD KQK/ AYSLLD KQK/ AYSLLD FQK/	Pol	III (dNTP bit 640 GFTGTVNNCMI GFTGTVNNCMI GFTGTVNNCMI GFTGTVNNCMI GFTGTVNNCMI GFTGTVNNCMI GFTGTVNNCMI	nding site) 650 	660 STGRDMI RR STGRDMI RR STGRDMI RR STGRDMI RR STGRDMI RR STGRDMI RR STGRDMI RR	670 TKEYAEKEHGC TKEYAEKEHGC TKEYAEKEHGC TKEYAEKEHGC TKEYAEKEHGC TKEYAECEHGC TKEYAECEHGC TKEYAECEHGC	IVN IVN IVN IVN IVN IVN IVN IVN
EsV-1 genome Esil 1 Efas 1 Efas 2 Esil genome provirus Hinc_1 Mcla 1 Flit 1 Flex 8	580 	Pol 590  RSILCELEEN RSILCELEEN RSILCELEEN RSILCELEEN KSILCELEEN KSILCELEEN KSILCELEEN KSILCELEEN	VI 600 LIGEERKATKI LIGEERKATKI LIGEERKATKI LIGEERKATKI LIGEERKATKI LIGEERKATKI LIGEERKATKI LIGEERKATKI	610 	620 AYSLLDSKQK/ AYSLLDSKQK/ AYSLLDSKQK/ AYSLLDSKQK/ AYSLLDSKQK/ AYSLLDSKQK/ AYSLLDSKQK/ AYSLLDSKQK/ AYSLLDS?QK/ RYSLLNSKQL/	Pol 630 QKVTMINSIY QKVTMINSIY AQKVTMINSIY AQKVTMINSIY AQKVTMINSIY AQKVTMINSIY AQKVTMINSIY	III (dNTP bit 640 	nding site) 650 	660 STGRDMI CR STGRDMI CR STGRDMI CR STGRDMI CR STGRDMI CR STGRDMI CR STGRDMI CR STGRDMI CR STGRDMI CR STGRDMI CR	670 TKE YAEKEHGC TKE YAEKEHGC TKE YAEKEHGC TKE YAEKEHGC TKE YAEKEHGC TKE YAEAETEHGC TKE YAEAETEHGC TKE YAEAEAEGC TSD YV-KSLGA	NVI NVI NVI NVI NVI NVI NVI NVI
EsV-1 genome Esil 1 Efas 1 Efas 2 Esil genome provirus Hinc_1 Mcla 1 Plit 1 Flex 8 FsV-158 genome	580 ISDGVSVTFAHRPGS ISDGVSVTFAHRPGS ISDEALVTFAHRPGS ISDGVSVTFAHRSGS ISDGVSVTFAHRSGS VSEGVATFAHRPGS VAGGVTATFAHRPGS ISPGVETFAYRAGD ISPGVETFAYRACD	Pol	VI 600 LLGERRATKH LLGERRATKH LLGERRATKH LLGERRATKH LLGERRATKH LLGERRATKH LLGERRATKH LLGERRATKH LLGERRATKH LLGERRATKH	610 	620 AYSLID KQKI AYSLID KQKI AYSLID KQKI AYSLID KQKI AYSLID KQKI AYSLID KQKI AYSLID RQKI AYSLID RQKI KUQLIN KQLI	Pol 630 AQKVTMINSIY AQKVTMINSIY AQKVTMINSIY AQKVTMINSIY AQKVTMINSIY AQKVTMINSIY AQKVTMINSIY AQKVTMINSIY AQKVTMINSIY	III (dNTP bit 640 GFTGTVNNCHI GFTGTVNNCHI GFTGTVNNCHI GFTGTVNNCHI GFTGTVNNCHI GFTGTVNNCHI GFTGTVNNCHI GFTGTVNNCHI	nding site) 650 	660 STGRDMI KR STGRDMI KR STGRDMI KR STGRDMI KR STGRDMI KR STGRDMI KR CTGRDMI KA CTGRDMI KA	670 TREYARKEHGC TREYARKEHGC TREYARKEHGC TREYARKEHGC TREYARKEHGC TREYAREHGC TREYARDEHGC TREYARDEHGC TREYARDEHGC TADFTREMGG	NVI NVI NVI NVI NVI NVI NVI EVV TVI
EsV-1 genome Esil 1 Efas 1 Efas 2 Esil genome provirus Hinc_1 Mcla 1 Plit 1 Flex 8 Fiv-158 genome FirrV-1 genome	580 III ISDGVSVTFAHRPGS ISDGVSVTFAHRPGS ISDGVSVTFAHRGS ISDGVSVTFAHRGS VSEGVIATFAHRGS VSEGVIATFAHRGS TSPGRVETFARRGD ISPGVVETFARRCR VAGVVETFARRPGR	Pol 590 RSILCELEEN RSILCELEEN RSILCELEEN RSILCELEEN KSILCELEEN KSILCELEEN KSILCELEEN KSILSQIEE SILSQIEE SILSQIEE	VI 600 LIGEERKATKI LIGEERKATKI LIGEERKATKI LIGEERKATKI LIGEERKATKI LIGEERKATKI LIGEARRATKI LIGEARRITKI LIGEARRITKI	610 	620 AYSLLD KQKV AYSLLD KQKV AYSLLD KQKV AYSLLD KQKV AYSLLD KQKV AYSLLD KQKV AYSLLD KQKV AYSLLD KQLV KQLIN KQLI LLQLIN KQLI	Pol 630 AQKVTMINSTY AQKVTMINSTY AQKVTMINSTY AQKVTMINSTY AQKVTMINSTY AQKVTMINSTY AQKVTMINSTY AQKVTMINSTY AQKVTMINSTY AQKVTMINSTY AQKVTMINSTY	III (dNTP bit 640 GFTGTVNICHI GFTGTVNICHI GFTGTVNICHI GFTGTVNICHI GFTGTVNICHI GFTGTVNICHI GFTGTVNICHI GFTGTVNICHI GFTGTVNICHI GFTGTV-NICHI GFCGTV-NICHI GFCGTV-NICHI	nding site) 650 	660 STGRDMI CR STGRDMI CR STGRDMI CR STGRDMI CR STGRDMI CR STGRDMI CR CTGRDMI CR CTGRDMI CR CTGRCMI CR CTGRCMI CR	670 TREYARKEHGC TREYARKEHGC TREYARKEHGC TREYARKEHGC TREYAREHGC TREYAREHGC TREYAREHGC TREYAREHGC TSDYV-KSLGA TADFTREMGG TADFTREMGG	NVI NVI NVI NVI NVI NVI EVV TVI VVI
EsV-1 genome Esil 1 Efas 1 Efas 2 Esil genome provirus Hinc_1 Mcla 1 Plit 1 Flex 8 FSV-158 genome FirrV-1 genome Firr 1 Eirr 2	580 IIISDGVSVTFAHRPGS ISDGVSVTFAHRPGS ISDGVSVTFAHRPGS ISDGVSVTFAHRPGS VAGVSVTFAHRPGS VAGVTATFAHRPGS TSPGRVETFATRGH VSAGVYETFATRGH VSAGVYETFATRGH VSAGVYETFATRGH	Pol 590  RSILCELEEN RSILCELEEN RSILCELEEN RSILCELEEN KSILCELEEN KSILCELEEN KSILSQIEE PSILAMIEEI PSILAMIEEI	VI 600 LIGEERKATKH LIGEERKATKH LIGEERKATKH LIGEERKATKH LIGEERKATKH LIGEERKATKH JIGEERKATKH DIGEARKITKH DIGEARKITKH	610 	620 AYSLLD KQKI AYSLLD KQKI AYSLLD KQKI AYSLLD KQKI AYSLLD KQKI AYSLLD KQKI AYSLLD SQKI AYSLLD SQKI KUQLIN KQLI LLQLIN KQLI LLQLIN KQLI	Pol 630 	III (dNTP bit 640 GFTGTVNICHI GFTGTVNICHI GFTGTVNICHI GFTGTVNICHI GFTGTVNICHI GFTGTVNICHI GFTGTVNICHI GFTGTVNICHI GFCGTV-NGCI GFCGTV-NGCI GFCGTV-NGCI GFCGTV-NGCI	nding site) 650 	660 STGRDMI CR STGRDMI CR STGRDMI CR STGRDMI CR STGRDMI CR STGRDMI CR STGRDMI CR CTGRSMI CT CTGRSMI CR CTGRSMI CR CTGRSMI CR CTGRSMI CR CTGRSMI CR CTGRSMI CR	670 TKEYAEKEHGC TKEYAEKEHGC TKEYAEKEHGC TKEYAEKEHGC TKEYAEKEHGC TKEYAEKEHGC TKEYAEKEHGC TKEYAEKEHGC TKEYAEKEHGC TADFIRTEMGG TADFIRTEMGG	NVI NVI NVI NVI NVI NVI NVI VVI VVI VVI
EsV-1 genome Esil 1 Efas 1 Efas 2 Esil genome provirus Hinc_1 Mcla 1 Plit 1 Flex 8 FsV-158 genome FirrV-1 genome Firr 2 Firr 2 Elor 2	580 	Pol 590  RSILCELEER RSILCELEER RSILCELEER KSILCELEER KSILCELEER KSILCELEER KSILSIEEE PILAMIEEI PILAMIEEI PILAMIEEI	VI 600 LICEERKATKH LICEERKATKH LICEERKATKH LICEERKATKH LICEERKATKH LICEERKATKH LICEERKATKH DICEARRATKH DICEARRATKH DICEARRATKH DICEARRATKH DICEARRATKH DICEARRATKH DICEARRATKH	610 	620 AYSLLD KQKI AYSLLD KQKI AYSLLD KQKI AYSLLD KQKI AYSLLD KQKI AYSLLD KQKI AYSLLD FQKI KUQLIN KQLI LUQLIN KQLI LUQLIN KQLI LUQLIN KQLI	Pol 630 	III (dNTP bit 640 GFTGTVINCHI GFTGTVINCHI GFTGTVINCHI GFTGTVINCHI GFTGTVINCHI GFTGTVINCHI GFTGTVINCHI GFTGTVINCHI GFCGTV-NCHI GFCGTV-RCCI GFCGTV-RCCI GFCGTV-RCCI GFCGTV-RCCI	DISTINCT STATES	660 STGRDMI CR STGRDMI CR STGRDMI CR STGRDMI CR STGRDMI CR STGRDMI CR STGRDMI CR CTGRDMI CR CTGRDMI CR CTGRDMI CR CTGRDMI CR CTGRDMI CR CTGRDMI CR CTGRDMI CR CTGRDMI CR	670 TREYARKEHGC TREYARKEHGC TREYARKEHGC TREYARKEHGC TREYAREHGC TREYAREHGC TREYAREHGC TREYAREHGC TREYARDHG TADFINDKG TADFINDKG TADFINDKG TADFINDKG	NVI NVI NVI NVI NVI NVI NVI VVI VVI VVI
EsV-1 genome Esil 1 Efas 1 Efas 2 Esil genome provirus Hinc 1 Mola 1 Plit 1 Flex 8 FsV-158 genome FirrV-1 genome Firr 1 Firr 2 Flex 2 Flex 7	580 IISDEGVSVTFAHRPGS ISDEGVSVTFAHRPGS ISDEATVTFAHRPGS ISDEGVSVTFAHRPGS VSEGVATFAHRPGS VAGGVATFAHRPGS VAGGVETFARRPGR VSAGVYETFARRPGR VSAGVYETFARRPGR VNGVFETFARRPGR VTPGVFETFGRPGC	Pol 590  RSILCELEER RSILCELEER RSILCELEER KSILCELEER KSILCELEER KSILSAIEEI PSILAMIEEI PSILAMIEEI KSILSAIEEI KSILSMIEEI KSILSMIEEI	VI 600 LICEERKATKI LICEERKATKI LICEERKATKI LICEERKATKI LICEERKATKI LICEERKATKI DICEARKATKI DICEARKITKI DICEARKITKI DICEARKITKI DICEARKITKI	610 	620 AYSLLDEKOK AYSLLDEKOK AYSLLDEKOK AYSLLDEKOK AYSLLDEKOK AYSLLDEKOK AYSLLDEKOK KLOLINEKOL KLOLINEKOL AYOLINEKOL AYOLINEKOL	Pol 630 AQKVTMINSTY AQKVTMINSTY AQKVTMINSTY AQKVTMINSTY AQKVTMINSTY AQKVTMINSTY AQKVTMINSTY AQKVTMINSTY AQKITMINSTY AQKITMINSTY AQKUTMINSTY	III (dNTP bit 640 GFTGTVINGHI GFTGTVINGHI GFTGTVINGHI GFTGTVINGHI GFTGTVINGHI GFTGTVINGHI GFTGTVINGHI GFTGTVINGHI GFTGTV-NGCI GFCGTV-RGCI GFCGTV-RGCI GFCGTV-RGCI GFCGTV-RGVI	nding site) 650 	660 STGRDMI R STGRDMI R STGRDMI R STGRDMI R STGRDMI R STGRDMI R STGRDMI R STGRDMI R CTGRDMI R ATGRMI R CTGRSMI R CATGRAMI R CATGRAMI R	670 TREYARKEHGC TREYARKEHGC TREYARKEHGC TREYARKEHGC TREYAREHGC TREYAREHGC TREYAREHGC TREYAREHGC TREYARRANGC TADFIRNDKG TADFIRNDKG TADFIRNDKG TADFIRNDKG TADFIRNDKG	NVI NVI NVI NVI NVI NVI VVI VVI VVI VVI
EsV-1 genome Esil 1 Efas 1 Efas 2 Esil genome provirus Hinc 1 Mcla 1 Plit 1 Flex 8 FsV-158 genome FirrV-1 genome Firr 2 Flex 2 Flex 7 Flex 6	580 III ISDGVSVTFAHRPGS ISDGVSVTFAHRPGS ISDGVSVTFAHRPGS ISDGVSVTFAHRPGS VSEGVJATFAHRPGS VSEGVJATFAHRPGS VSEGVJATFAHRPGS VSAGVJETFARRPGR VSAGVJETFARRPGR VSAGVJETFARRPGR VNKGVFETFARRPGR VNGVFETFARRPGR VTPGVFETFGRRPGQ VTPGVFETFGRRPGQ	Pol 590  R\$ILCELEER R\$ILCELEER R\$ILCELEER R\$ILCELEER K\$ILCELEER K\$ILCELEER K\$ILSIEE P\$ILAMIEE P\$ILAMIEE P\$ILAMIEE R\$ILSMIEE K\$ILSMIEE K\$ILSMIEE	VI 600 LIGEERKATKI LIGEERKATKI LIGEERKATKI LIGEERKATKI LIGEERKATKI LIGEARKATKI LIGEARKATKI DIGEARKITKI DIGEARKITKI DIGEARKITKI DIGEARKITKI DIGEARKITKI	610 	620 AYSLLD KQK/ AYSLLD KQK/ AYSLLD KQK/ AYSLLD KQK/ AYSLLD KQK/ AYSLLD KQK/ AYSLLD KQK/ AYSLLD FQK/ AYSLLD FQK/ LLQLIN KQL/ VLQLIN KQL/ VLQLIN KQL/ VLQLIN KQL/ VLQLIN KQL/ VLQLIN KQL/ VLQLIN KQL/	Pol 630 	III (dNTP bit 640 (GFTGTVNICMI GFTGTVNICMI GFTGTVNICMI GFTGTVNICMI GFTGTVNICMI GFTGTVNICMI GFTGTVNICMI GFCGTV-NGCI GFCGTV-NGCI GFCGTV-RGCI GFCGTV-RGCI GFCGTV-RGCI GFCGTA-KGCI GFCGTA-KGCI	nding site) 650 	660 STGRDMI R STGRDMI R STGRDMI R STGRDMI R STGRDMI R STGRDMI R CTGRDMI R CTGRDMI R CTGRDMI R CTGRDMI R CTGRCMI R R TGRRMI R ATGRMI R SATGRAMI K SATGRAMI K	670 TREYARKEHGC TREYARKEHGC TREYARKEHGC TREYARKEHGC TREYAREHGC TREYARREHGC TREYARREHGC TREYARREHGC TADFIRIDMGG TADFIRIDMGG TANFIRRDMGG TANFIRRDMGG TANFIRRDMGG	NVI NVI NVI NVI NVI NVI VVI VVI IVI IVI
EsV-1 genome Esil 1 Efas 1 Efas 2 Esil genome provirus Hinc_1 Mcla 1 Plit 1 Flex 8 FsV-158 genome FirrV-1 genome Firr 2 Flex 2 Flex 7 Flex 5	580 	Pol 590  RSILCELEEN RSILCELEEN RSILCELEEN RSILCELEEN KSILCELEEN KSILSQIEEZ KSILSQIEEZ RSILSQIEEZ RSILSMIEEI RSILSMIEEI RSILSMIEEI RSILSMIEEI	VI 600 ILGEERKATKH ILGEERKATKH ILGEERKATKH ILGEERKATKH ILGEERKATKH ILGEERKATKH ILGEARKATKH ILGEARKLTKH ILGEARKLTKKH ILGNARTKTKH ILGNARTKTKH ILGNARTKTKH ILGNARTKTKH	610 	620 AYSLLD KQKA AYSLLD KQKA AYSLLD KQKA AYSLLD KQKA AYSLLD KQKA AYSLLD KQKA AYSLLD KQKA AYSLLD SQKA AYSLLD SQKA AYSLLN KQLA LLQLLN KQLA AYQLLN KQLA VIQLLN KQLA VIQLLN KQLA VIQLLN KQLA IQLLN KQLA IQLLN KQLA	Pol 630 	III (dNTP bit 640 GFTGTVINCHI GFTGTVINCHI GFTGTVINCHI GFTGTVINCHI GFTGTVINCHI GFTGTVINCHI GFTGTVINCHI GFTGTVINCHI GFCGTV-RCCI GFCGTV-RCCI GFCGTV-RCCI GFCGTV-RCCI GFCGTV-RCCI GFCGTA-KCCI GFCGTA-KCCI GFCGTA-KCCI	nding site) 650 	660 STGRDMI RR STGRDMI RR STGRDMI RR STGRDMI RR STGRDMI RR STGRDMI RR STGRDMI RR CTGRSMI SR CTGRSMI SR ATGRPMI RR ATGRPMI RR ATGRAMI NK ATGRAMI NK ATGRAMI NK ATGRAMI NK	670 TKEYAEKEHGC TKEYAEKEHGC TKEYAEKEHGC TKEYAEKEHGC TKEYAEKEHGC TKEYAEGEHGC TKEYAEGEHGC TKEYAEGEHGC TKEYAEGEHGC TADFIRTMGG TADFIRTMGG TADFIRTMGG TANFIRRDMGG TANFIRRDMGG TAFFIRDMGG	NVI NVI NVI NVI NVI NVI NVI NVI NVI VVI V
EsV-1 genome Esil 1 Efas 1 Efas 2 Esil genome provirus Hinc_1 Mcla 1 Plit 1 Flex 8 FsV-158 genome Firrv-1 genome Firr 2 Flex 2 Flex 2 Flex 6 Flex 5 Flex 3	580 	Pol 590  RSILCELEER RSILCELEER RSILCELEER KSILCELEER KSILCELEER KSILCELEER KSILSAIEE RSILSAIEE RSILSAIEE KSILSAIEE KSILSMIEEI KSILSMIEEI KSILSMIEEI KSILSMIEEI	VI 600 LIGEERKATKH LIGEERKATKH LIGEERKATKH LIGEERKATKH LIGEERKATKH LIGEERKATKJ LIGEARKATKJ LIGEARKLTKH DIGEARKLTKH DIGEARKLTKH DIGEARKLTKH DIGEARKLTKH DIGEARKLTKH DIGEARKTKH	610 	620 AYSLID KQK/ AYSLID KQK/ AYSLID KQK/ AYSLID KQK/ AYSLID KQK/ AYSLID KQK/ AYSLID KQK/ AYSLID KQL/ KULLIN KQL/ AYQLIN KQL/ AYQLIN KQL/ AYQLIN KQL/ IIQLIN KQL/ IIQLIN KQL/ IIQLIN KQL/ IIQLIN KQL/ IIQLIN KQL/ IIQLIN KQL/ IIQLIN KQL/	Pol 630 AQKVTMINSTY AQKVTMINSTY AQKVTMINSTY AQKVTMINSTY AQKVTMINSTY AQKVTMINSTY AQKVTMINSTY AQKVTMINSTY AQKITMINSTY AQKITMINSTY AQKVTMINSTY AQKVTMINSTY AQKVTMINSTY	III (dNTP bit 640 GFTGTVNNCHI GFTGTVNNCHI GFTGTVNNCHI GFTGTVNNCHI GFTGTVNNCHI GFTGTVNNCHI GFTGTVNNCHI GFCGTV-NCCI GFCGTV-RCCI GFCGTV-RCCI GFCGTV-RCCI GFCGTA-KGCI GFCGTA-KGCI GFCGTA-KGCI GFCGTA-KGCI	nding site) 650 	660 STGRDMI R STGRDMI R STGRDMI R STGRDMI R STGRDMI R STGRDMI R STGRDMI R STGRDMI R STGRDMI R ATGRMI R ATGRMI R ATGRMI R ATGRAMI W ATGRAMI W ATGRAMI W	670 TREYAEKEHGC TREYAEKEHGC TREYAEKEHGC TREYAEKEHGC TREYAEKEHGC TREYAEKEHGC TREYAEKEHGC TREYAEKEHGC TREYAEKEHGC TREYAEKEHGC TADFIRDMGG TADFIRDMGG TAPFIREMGG TAPFIREMGG	NVI NVI NVI NVI NVI NVI TVI TVI TVI TVI TVI
EsV-1 genome Esil 1 Efas 1 Efas 2 Esil genome provirus Hinc_1 Mcla 1 Plit 1 Flex 8 FsV-158 genome FirrV-1 genome Firr 1 Firr 2 Flex 2 Flex 7 Flex 6 Flex 5 Flex 4	580 	Pol 590 RSILCELEEN RSILCELEEN RSILCELEEN RSILCELEEN KSILCELEEN KSILSQIEEE KSILSQIEEE KSILSMIEEI KSILSMIEEI KSILSMIEEI KSILSMIEEI KSILSMIEEI	VI 600 LLGEBRATKH LLGEBRKATKH LLGEBRKATKH LLGEBRKATKH LLGEBRRATKH LLGEBRRATKH LLGEARKLTKH DLGEARKLTKH DLGEARKLTKH DLGEARKLTKH DLGEARKLTKH DLGEARKLTKH DLGEARKLTKH DLGEARKLTKH DLGARKTTKJ DLGDARKTTKJ DLGDARKTTKJ	610 	620 AYSLLD KQK/ AYSLLD KQK/ AYSLLD KQK/ AYSLLD KQK/ AYSLLD KQK/ AYSLLD KQK/ AYSLLD KQL/ KYSLLN KQL/ KUQLLN KQL/ AYQLLN KQL/ KUQLLN KQL/ KUQLLN KQL/ KUQLLN KQL/ KUQLLN KQL/ KUQLLN KQL/	Pol 630 AQKVTMINSTY AQKVTMINSTY AQKVTMINSTY AQKVTMINSTY AQKVTMINSTY AQKVTMINSTY AQKVTMINSTY AQKITMINSTY AQKITMINSTY AQKVTMINSTY AQKVTMINSTY AQKVTMINSTY AQKVTMINSTY AQKVTMINSTY	III (dNTP bil 640 (GFTGTVNNCMI GFTGTVNNCMI GFTGTVNNCMI GFTGTVNNCMI GFTGTVNNCMI GFTGTVNNCMI GFTGTVNNCMI GFCGTV-NCCI GFCGTV-RCCI GFCGTV-RCCI GFCGTA-KGCI GFCGTA-KGCI GFCGTA-KGCI GFCGTA-KGCI GFCGTA-KGCI GFCGTA-KGCI GFCGTA-KGCI	nding site) 650 	660 STGRDMI CR STGRDMI CR STGRDMI CR STGRDMI CR STGRDMI CR STGRDMI CR STGRDMI CR CTGRDMI CR CTGRDMI CR CTGRCMI CR CTGRCMI CR CTGRCMI CR CTGRCMI CR CATGRAMI CR	670 TREYARKEHGC TREYARKEHGC TREYARKEHGC TREYARKEHGC TREYARKEHGC TREYAREHGC TREYARRAHGC TADFIREMGG TADFIREMGG TADFIREMGG TANFIRRMGG TANFIRRMGG TAFIRREMGG TAFIRREMGG TAFIRREMGG TAFIRREMGG TAFIRREMGG	NVI NVI NVI NVI IVI IVI IVI IVI IVI IVI
EsV-1 genome Esil 1 Efas 1 Efas 2 Esil genome provirus Hinc_1 Mcla 1 Plit 1 Flex 8 FSV-158 genome FirrV-1 genome Firr 2 Flex 2 Flex 7 Flex 6 Flex 5 Flex 3 Flex 4 Flex 1	580 III ISDGVSVTFAHRPGS ISDGVSVTFAHRPGS ISDGVSVTFAHRPGS ISDGVSVTFAHRPGS ISDGVSVTFAHRPGS VSEGVJATFAHRPGS VSEGVJATFAHRPGS VSGVYETFARRGR VSAGVYETFARRGR VSAGVYETFARRGR VNKGVFETFARRCR VNGVFETFARRCR VTPGVFETFCRRCQ ISSGVFETFCRRCQ ISSGVFETFCRRCQ ISSGVFETFCRRCQ ISSGVFETFCRRCQ	POI 590  RSILCELEER RSILCELEER RSILCELEER RSILCELEER KSILCELEER KSILCELEER KSILSQIEE DILAMIEEI POILAMIEEI KSILSMIEEI KSILSMIEEI KSILSMIEEI KSILSMIEEI KSILSMIEEI	VI 600 ILGEERKATKI ILGEERKATKI ILGEERKATKI ILGEERKATKI ILGEERKATKI ILGEERKATKI ILGEARKATKI DLGEARKLTKI DLGAARKITKI DLGDARKTTKI DLGDARKTTKI ILGDARKTTKI	610 	620 AYSLLD KQK/ AYSLLD KQK/ AYSLLD KQK/ AYSLLD KQK/ AYSLLD KQK/ AYSLLD KQK/ AYSLLD KQK/ AYSLLD FQK/ AYSLLD FQK/ AYSLLN KQL/ LLQLIN KQL/ UQLLN KQL/ UQLIN KQL/ MLQLIN KQL/ MLQLIN KQL/	Pol 630 QKVTMINSIY QKVTMINSIY AQKVTMINSIY AQKVTMINSIY AQKVTMINSIY AQKVTMINSIY AQKVTMINSIY AQKVTMINSIY AQKVTMINSIY AQKVTMINSIY AQKVTMINSIY AQKVTMINSIY AQKVTMINSIY AQKVTMINSIY AQKVTMINSIY AQKVTMINSIY AQKVTMINSIY	III (dNTP bil 640 (GFTGTVNICHI GFTGTVNICHI GFTGTVNICHI GFTGTVNICHI GFTGTVNICHI GFTGTVNICHI GFTGTVNICHI GFTGTVNICHI GFCGTV-NGCI GFCGTV-RGCI GFCGTV-RGCI GFCGTA-KGCI GFCGTA-KGCI GFCGTA-KGCI GFCGTA-RGCI GFCGTA-RGCI GFCGTV-RGCI	hding site) 650 	660 STGRDMI CR STGRDMI CR STGRDMI CR STGRDMI CR STGRDMI CR STGRDMI CR STGRDMI CR CTGRDMI CR CTGRDMI CR CTGRCMI CR C	670 	NAI NAI NAI NAI NAI NAI NAI NAI NAI NAI

**Figure 2. Partial nucleotide alignment of cloned fragment of the viral DNA polymerase gene.** Numbers refer to amino acid position in the complete EsV-1 DNA polymerase gene taken from Delaroque *et al.* 2001 [20] (GenBank accession number NC\_002687.1). Boxed regions indicate conserved polymerase domains [30]. \* indicates conserved positions between all sequences,  $\blacktriangle$  shows where the Flex 8 variant shares an amino acid with the larger viruses of sub-group A,  $\blacktriangledown$  shows where the Flex 8 variant shares an amino acid with the smaller genomed viruses of sub-group B. doi:10.1371/journal.pone.0086040.g002

known to be important for genome stability [19]. Daee et al. found that a single amino acid mutation was associated with extreme rates of spontaneous mutation in yeast. Three amino acid polymorphisms in this region (Figure 2, positions 618, 621, 625) could be a contributing factor to the large number of variants observed amongst the Feldmannia viruses. The second key observation is that Flex 8, at the base of the sub-group B clades in all the phylogenetic trees, is probably the progenitor virus to the Feldmannia sub-group B viruses. This, therefore, gives us a unique insight into the emergence of a new phaeovirus sub-group, which is likely to be a result of the genome reduction of an ancestral member from sub-group A. Both FirrV-1 and FsV-158 [15] show the loss of the DNA proofreading exonuclease gene (EsV-126) known to be present in EsV-1 [20]. We therefore hypothesize that these genomic modifications could have resulted in the key life strategy shift, thereby utilizing the high mutation rates more associated with acute infections. When and how this happened is unclear; however, the expansion of localised genomic regions in poxviruses, causing gene duplications and mutations have been proposed to be a response to overcoming changing immune responses after a host switch [21]. Whilst gene duplications have not yet been discovered in the phaeoviruses sequenced thus far, these expansions are usually followed by a rapid gene reduction in order to minimise the burden of replicating and enlarged genome, therefore a similar mechanism may also be involved here.

A pairwise analysis of the evolutionary divergence in nucleotide sequences within the various groups of phycodnaviruses (Figure 5) illustrates the shift by sub-group B to a genome characteristic of an *r*-like evolutionary strategy. Sub-group B has a median nucleotide divergence of 29.3% in the DNApol gene fragment, comparable to that of the other *r*- selected lytic phycodnavirus groups (24.3–47.9%). Sub-group A has maintained the classic K – selection life strategy with a much lower divergence of 14.9%.

There have been several studies that reported on the host specificity of phaeoviruses. EsV-1 can successfully infect *Kuckuckia kylinii* (Cardinal) producing virions infectious to the original host [22]. Other cross-species infections do not produce infectious virions although the virus does induce symptom-like deformities in the host, for example EsV-1 in *F. simplex* [22], or EfasV in *E. siliculosus* [23] and *M. clavaeformis* [24]. This demonstrates not only the potential of phaeoviruses to jump between species but also that not all jumps result in successful infections. Another example of this unsuccessful jump can be seen by the presence of an inactive provirus in the *Ectocarpus* genome [4]. Here we show that the



**Figure 3. Maximum Likelihood analysis between variants of the phaeoviral sequences of DNA polymerase.** Single value node labels represent ML bootstrap values. Where nodes are labelled with two values, this indicates that both ML and Bayesian topologies agree (whole numbers represent ML bootstrap values, decimals indicate Bayesian posterior probability). Sub-group A viruses are labelled in blue, sub-group B viruses are red and the intermediate Flex virus variant is green. Bold values are those greater than 75% bootstrap or probability. doi:10.1371/journal.pone.0086040.g003

provirus appears to be more closely related to an *E. fasciculatus* variant than to EsV-1 (e.g. Figures 3 & 4). This suggests that an *E. fasciculatus* virus infected an *Ectocarpus* species more closely related to *E. siliculosus* [25]. This study also confirmed the life-history and morphometric data that the viruses infecting Efas, Mcla, Plit and Hinc do indeed belong in the phaeovirus group. Moreover, there is also a corresponding grouping which can be created based on genome sizes (see Table 1); the larger viral genomes from Esil, Efas, Plit, Mcla and Hinc (240–336 kb) fall within sub-group A and the smaller viruses from Firr, Flex and *Feldmannia* sp [15] (158–220 kb) into sub-group B.

This study provides the first example of an emergent virus system retaining a Persistent life strategy, but exploiting an Acute strategist's high genomic mutation rate. Moreover, unlike current reports on how emerging acute diseases develop where cryptic persistent viruses cross species boundaries (e.g. HIV [26], H5N1 [27] and DWV [28]), which can have catastrophic consequences for new host survival, this study suggests a very different scenario of one in which the integration and diversification of Persistent viruses has been stably maintained over a long period of time. Similarly, due to their evolutionary link to animal viruses this infection strategy is likely to also occur in these systems, and further studies in this field may help our understanding of the spread of new emergent diseases.

#### **Materials and Methods**

#### Isolates & culture conditions

See Table 1 for a list of the phaeovirus-infected cultures used in this study. Each strain was cultured in a 40 ml petri dish at  $15^{\circ}$ C, 16:8 light-dark cycle, approximately 100 µmol photons m<sup>-2</sup> s<sup>-1</sup>. The Western Channel Observatory (www. westernchannelobservatory.org.uk) is an oceanographic time-series and marine biodiversity reference site in the Western English Channel. In situ measurements are undertaken weekly at coastal station L4 (source of water for our study) and fortnightly at open



**Figure 4. Maximum Likelihood analysis between variants of the phaeoviral sequences of concatenations of DNApol and MCP.** Variants are labelled according to DNApol identifier initially, followed by the MCP variant number in brackets. In order to slightly reduce the number of combinations of sequences, where individual gene phylogenies show a clear separation of individual variants, these are concatenated together and excluded from the other combinations. Single value node labels represent ML bootstrap values. Where nodes are labelled with two values, this indicates that both ML and Bayesian topologies agree (whole numbers represent ML bootstrap values, decimals indicate Bayesian posterior probability). Sub-group A viruses are labelled in blue, sub-group B viruses are red and the intermediate Flex virus variant is green. Bold values are those greater than 75% bootstrap or probability. doi:10.1371/journal.pone.0086040.q004

shelf station E1 using the research vessels of the Plymouth Marine Laboratory and the Marine Biological Association. THE DATA POLICY of the NERC National Capability funded Western Channel Observatory is to make the data freely available at the point of delivery. Culture medium was filtered (30 kDa) natural sea water from the L4 sampling station close to the Eddystone Lighthouse near Plymouth, enriched with Provasoli's enrichment [29]. Sub-culturing into a new dish with fresh media was carried out every 14 days, when the cultures were pulled apart using forceps to separate out filaments in order to encourage production of zoidangia and virions.

#### DNA extraction method

 $50\mathchar`-200$  mg wet weight fresh algal material was transferred to an Eppendorf tube, frozen in liquid nitrogen and ground using

Eppendorf grinders with 10 µl saturated  $\leq$ 106 microns acid washed glass bead solution before proceeding with the Qiagen DNeasy protocol for Genomic DNA purification from cultured animal cells, starting with the proteinase K treatment. 40 µl proteinase K and 200 µl Buffer AL were added to the sample and incubated at 56°C for 30 minutes, before centrifuging for 2 minutes at maximum speed to separate out the beads. 200 µl ethanol was added to the resulting supernatant, vortexed and pipetted onto the spin column, to proceed with the first centrifugation step. For the final step, DNA was eluted using 100 µl water, instead of 200 µl in order to obtain a more concentrated sample.



**Figure 5. Box and whiskers plot of evolutionary divergence between nucleotide sequences of the DNApol.** Identical sequences were not included more than once. The box represents the interquartile range which shows the middle 50% of the data, the bottom line being the first quartile, the middle line being the median and the upper line being the third quartile. The whiskers represent the maximum (or minimum) data point up to 1.5 times the box height above (or below) the top (or bottom) of the box. Outliers beyond the whiskers are shown as a \*. Phaeovirus sub-groups are as shown in Fig. 3 & 4, with Flex 8 being included in sub-group B. Chloroviruses consist of thirteen viral isolates from *Paramecium bursaria Chlorella* (AF344202, AF344203, AF344211, AF344215, AF344215, AF344230, AF344231, AF344235, AF344238, AF344239, M86837, U32985) and one from *Acanthocystis turfacea Chlorella* (AY971002). The other phycodnaviruses group consists of three viral isolates from *Emiliania huxleyi* (AF453961, AF453867, AF472534), three from *Micromonas pusilla* (U32975, U32982, U32976), five from *Ostreococcus tauri* (FJ67503, FJ884758, FJ884773, FJ884773, FJ884776), two from *Ostreococcus lucimarinus* (GQ412090, GQ412099), six from *Phaoecystis globosa* (A345136-AY345140, DQ401030), one from *Chrysochromulina brevifilum* (U32983), one from *Chrysochromulina ericina* (EU006632) and one from *Heterosigma akashiwo* (AB194136).

doi:10.1371/journal.pone.0086040.g005

#### PCR, cloning & Sequencing

Degenerate primers were designed for three active viral genes (DNA polymerase (GRGGNCAGCAGATYAAGTG forward, GARTCCGTRTCSCCRTA reverse), helicase (GTGGCAGGT-SATYCCYTTC forward, GTTKCCGGCCATGATYCC reverse) and major capsid protein (MCP) (CVGCGTACTGGGT-GAACGC forward, AGTACTTGTTGAACCAGAACGG reverse)) against a consensus of published sequences from EsV-1, FirrV-1, FsV-158 and the provirus from the sequenced Ectocarpus genome. Degenerate PCR was carried out using Promega GoTaq<sup>®</sup> Flexi DNA polymerase kit, with an addition of 0.8 mg/ml bovine serum antigen (BSA). Cycling conditions were 95°C for 5 minutes, followed by 35 cycles of 95°C for 1 minute, a 30 second annealing step, an extension step at 72°C, and a final elongation step at 72°C for 10 minutes. Oligonucleotide and magnesium concentrations, annealing temperatures and extension times varied for each gene: DNApol required 1.25 mM MgCl<sup>2</sup>, 4 pmol/µl oligonucleotides, 50°C annealing temperature and 10 second extension time, MCP required 1.5 mM MgCl<sup>2</sup>, 8 pmol/ $\mu$ l oligonucleotides, 55°C annealing temperature and 30 second extension time and helicase required 1.5 mM MgCl<sup>2</sup>, 8 pmol/ $\mu$ l oligonucleotides, 55°C annealing temperature and 10 second extension time. Post-PCR samples were run on a 2% agarose gel at 80 V to achieve maximum separation between the bands. Samples with more than one product were purified by gel extraction; the band of the correct size was cut out of the gel and purified using the Qiaex II® Gel Extraction Kit. Samples with clean bands were purified using GenElute<sup>TM</sup> PCR Clean-Up Kit from Sigma. Purified PCR product was cloned into pCR<sup>®</sup>2.1, incubated overnight at 15°C before storing at  $-20^{\circ}$ C until used. 4 µl ligation mixture was added to 0.2 ml competent cells and mixed. The cells were then incubated on ice for 40 minutes, heat shocked at 42°C for 2 minutes and returned to the ice for 5 minutes. 0.7 ml prewarmed LB medium was added to the cells which were then incubated at 37°C for one hour. The cells were concentrated by spinning at 8000 g for 5 minutes, removing 0.5 ml supernatant, and re-suspended gently with a pipette before being plated out onto LB agar plates containing 5 µg/ml ampicillin, with 40 µl of 20 X-gal spread on each plate. Plates were incubated overnight at 37°C.

Single cloned colonies were picked from agar plates into individual 0.2 ml tubes containing 5  $\mu$ l molecular grade water and heated to 95°C for 5 minutes to denature the cells before adding 10  $\mu$ l 5× buffer, 5  $\mu$ l 25 mM MgCl<sub>2</sub>, 5  $\mu$ l 2.5 mM dNTPs, 2  $\mu$ l each of 10 pmol/ $\mu$ l M13 forward and reverse primers, 0.2  $\mu$ l Taq polymerase, 20.8  $\mu$ l molecular grade H<sub>2</sub>O. Cycling conditions consisted of 30 cycles of 95°C for 45 seconds, 56°C for 45 seconds and 72°C for 45 seconds, followed by a final extension step of 72°C for 5 minutes.

PCR products were purified using the Qiaex II<sup>®</sup> Gel Extraction Kit and then sequenced using the BigDye<sup>®</sup> Terminator v3.1. The mix consisted of 3.5 µl 5× BigDye buffer, 1 µl Ready Reaction Mix, 2 µl template (6–14 ng µl<sup>-1</sup> concentration), 1 µl primers (either M13 forward or reverse) at a concentration of 3.2 pmol µl<sup>-1</sup> and 12.5 µl dH<sub>2</sub>O. Cycling conditions were 95°C for 2 minutes, followed by 30 cycles of 95°C for 30 seconds, 50°C for 30 seconds, 72°C for 30 seconds, then a final elongation at 72°C for 5 minutes. Sequenced reactions were precipitated by adding 5 µl 125 mM EDTA and 65  $\mu$ l cold 100% ethanol and incubated in the dark at room temperature for 15 minutes. They were then spun for 30 minutes at 2200 g, the supernatant removed and the pellet washed with 60  $\mu$ l cold 70% ethanol, and spun for a further 15 minutes at 2200 g. The supernatant was removed again and the pellet air dried. Sanger sequencing was carried out by Source Bioscience in Cambridge. Sequences were submitted to the European Nucleotide Archive with accession numbers (HG003317–HG003355).

#### Phylogeny

Bayesian analysis of phylogenetic trees was carried out using MrBayes v3.2.1, running the analysis until the standard deviation of split frequencies reached <0.01 and the number of generations was >100 000. Maximum Likelihood analysis was carried out using MEGA5.1 WAG model with 500 bootstrap replications and the Nearest-Neighbour-Interchange heuristic method. DNApol and MCP sequences were combined in all possible combinations (Table 1) in order to create concatenations which were used to create Figure 4.

#### Distance analysis

Nucleotide sequences were obtained for the various groups of phycodnaviruses that have been sequenced to data by carrying out

#### References

- Roossinck MJ (2010) Lifestyles of plant viruses. Philosophical Transactions of the Royal Society B: Biological Sciences 365: 1899–1905.
- Villarreal LP, Defilippis VR, Gottlieb KA (2000) Acute and persistent viral life strategies and their relationship to emerging diseases. Virology 272: 1–6.
- Suttle CA (2007) Marine viruses major players in the global ecosystem. Nature reviews Microbiology 5: 801–812.
- Cock JM, Sterck L, Rouzé P, Scornet D, Allen AE, et al. (2010) The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. Nature 465: 617–621.
- Silberfeld T, Leigh JW, Verbruggen H, Cruaud C, De Reviers B, et al. (2010) A multi-locus time-calibrated phylogeny of the brown algae (Heterokonta, Ochrophyta, Phaeophyceae): Investigating the evolutionary nature of the "brown algal crown radiation". Molecular Phylogenetics and Evolution 56: 659–674.
- Bergh O, Borsheim KY, Bratbak G, Heldal M (1989) High abundance of viruses found in aquatic environments. Nature 340: 467–468.
- Proctor LM, Fuhrman JA (1990) Viral mortality of marine bacteria and cyanobacteria. Nature 343: 60–62.
- Suttle CA, Chan AM, Cottrell MT (1990) Infection of phytoplankton by viruses and reduction of primary productivity. Nature 347: 467–469.
- Fuhrman JA (1999) Marine viruses and their biogeochemical and ecological effects. Nature 399: 541–548.
- Brussaard CPD, Wilhelm SW, Thingstad F, Weinbauer MG, Bratbak G, et al. (2008) Global-scale processes with a nanoscale drive: the role of marine viruses. The ISME Journal 2: 575–578.
- Wilson WH, Schroeder DC, Allen MJ, Holden MTG, Parkhill J, et al. (2005) Complete genome sequence and lytic phase transcription profile of a Coccolithovirus. Science 309: 1090–1092.
- 12. Suttle CA (2005) Viruses in the sea. Nature 437: 356-361.
- Friess-Klebl AK, Knippers R, Müller DG (1994) Isolation and characterization of a DNA virus infecting *Feldmannia simplex* (Phaeophyceae). Journal of Phycology 30: 653–658.
- Schroeder D, Oke J, Malin G, Wilson W (2002) Coccolithovirus (Phycodnaviridae): characterisation of a new large dsDNA algal virus that infects *Emiliana huxleyi*. Archives of Virology 147: 1685–1698.
- Schroeder DC, Park Y, Yoon H-M, Lee YS, Kang SW, et al. (2009) Genomic analysis of the smallest giant virus - *Feldmannia* sp. virus 158. Virology 384: 223– 232.
- Schroeder DC (2011) Viruses of Seaweeds. Studies in Viral Ecology: Microbial and Botanical Host Systems 1: 205.

a BLAST search of known genome sequences from each group. The phaeovirus sequences obtained in this study were split into two subgroups according to their phylogenies as shown in Figures 3 & 4. Chloroviruses were considered together with the prasinoviruses, and the remaining viral groups (coccolithoviruses, prymnesioviruses, raphidoviruses) were considered together since they are all lytic viruses of stramenopiles or coccolithophores. Pairwise distances were computed using Mega 5.05.

#### Acknowledgments

We would like to thank Dieter Müller for kind donation of the infected ectocarpoid strains, Akira Peters for invaluable advice and discussion throughout this project. We would also like to thank David Dunigan & Christon Hurst for comments on this manuscript, as well as Matt Hall and Andrea Highfield from the Marine Biological Association and Rowena Stern from the Sir Alister Hardy Foundation for Ocean Science for their help and advice.

#### **Author Contributions**

Conceived and designed the experiments: KS DCS. Performed the experiments: KS KW CBellas SB DCS. Analyzed the data: KS KW CBellas MTB DCS. Contributed reagents/materials/analysis tools: CBrownlee DCS. Wrote the paper: KS MTB DCS.

- Ivey RG, Henry EC, Lee AM, Klepper L, Krueger SK, et al. (1996) A Feldmannia algal virus has two genome size-classes. Virology 220: 267–273.
- Delaroque N, Boland W, Müller DG, Knippers R (2003) Comparisons of two large phaeoviral genomes and evolutionary implications. Journal of Molecular Evolution 57: 613–622.
- Daee DL, Mertz TM, Shcherbakova PV (2010) A cancer-associated DNA polymerase δ variant modeled in yeast causes a catastrophic increase in genomic instability. Proceedings of the National Academy of Sciences of the United States of America 107: 157–162.
- Delaroque N, Müller DG, Bothe G, Pohl T, Knippers R, et al. (2001) The complete DNA sequence of the *Ectocarpus siliculosus* virus EsV-1 genome. Virology 287: 112–132.
- Elde NC, Child SJ, Eickbush MT, Kitzman JO, Rogers KS, et al. (2012) Poxviruses deploy genomic accordions to adapt rapidly against host antiviral defenses. Cell 150: 831–841.
- Müller DG, Parodi E (1993) Transfer of a marine DNA virus from *Eclocarpus* to *Feldmannia* (Ectocarpales, Phaeophyceae): aberrant symptoms and restitution of the host. Protoplasma 175: 121–125.
- Müller DG, Sengco M, Wolf S, Bráutigam M, Schmid CE, et al. (1996) Comparison of two DNA viruses infecting the marine brown algae *Ectocarpus* siliculosus and *E. fasciculatus*. Journal of General Virology 77: 2329–2333.
- 24. Maier I, Rometsch E, Wolf S, Kapp M, Müller DG, et al. (1997) Passage of a marine brown algal DNA virus from *Ectocarpus fasciculatus* (Ectocarpales, Phaeophyceae) to *Myriotrichia clavaeformis* (Dictyosiphonales, Phaeophyceae): infection symptoms and recovery. Journal of Phycology 33: 838–844.
- Peters AF, Serinde J, Cho GY, Scornet D, Hanyuda T, et al. (2010) Reinstatement of *Eclocarpus crouanionum* Thuret in Le Jolis as a third common species of *Eclocarpus* (Ectocarpales, Phaeophyceae) in Western Europe, and its phenology at Roscoff, Brittany. Phycological Research 58: 157–170.
- Gao F, Bailes E, Robertson DL, Chen Y, Rodenburg CM, et al. (1999) Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. Nature 397: 436–440.
- Claas ECJ, Osterhaus ADME, van Beek R, De Jong JC, Rimmelzwaan GF, et al. (1998) Human influenza A H5N1 virus related to a highly pathogenic avian influenza virus. The Lancet 351: 472–477.
- Martin SJ, Highfield AC, Brettell L, Villalobos EM, Budge GE, et al. (2012) Global honey bee viral landscape altered by a parasitic mite. Science 336: 1304– 1306.
- Starr RC, Zeikus JA (1993) The culture collection of algae at the University of Texas at Austin. J Phycol 29: 90–95.
- Preston BD, Albertson TM, Herr AJ (2010) DNA replication fidelity and cancer. Elsevier. pp. 281–293.