

## Supplementary figures:

### **Comprehensive analysis of single molecule sequencing-derived complete genome and whole transcriptome of *Hyposidra talaca* nuclear polyhedrosis virus**

Thong T. Nguyen<sup>1</sup>, Kushal Suryamohan<sup>1</sup>, Boney Kuriakose<sup>2,3</sup>, Vasantharajan Jankiraman<sup>1</sup>, Mike Reichelt<sup>4</sup>, Subhra Chaudhuri<sup>1</sup>, Joseph Guillory<sup>1</sup>, Neethu Divakaran<sup>2</sup>, Rabins PE<sup>2</sup>, Ridhi Goel<sup>2</sup>, Bhabesh Deka<sup>5</sup>, Suman Sarkar<sup>5</sup>, Preety Ekka<sup>5</sup>, Yu-Chih Tsai<sup>6</sup>, Derek Vargas<sup>1</sup>, Sam Santhosh<sup>7</sup>, Sangeetha Mohan<sup>7</sup>, Chen-Shan Chin<sup>6</sup>, Jonas Korfach<sup>6</sup>, George Thomas<sup>2,3</sup>, Azariah Babu<sup>5</sup> and Somasekar Seshagiri<sup>1\*</sup>

<sup>1</sup>Department of Molecular Biology, Genentech Inc., 1 DNA WAY, South San Francisco, CA, 94080, USA; <sup>2</sup>AgriGenome Labs Private Limited, 501, SCK01 Building, SmartCity Kochi, Infopark Road, Kakkanad, Kochi, Kerala, 682 042, India; <sup>3</sup>SciGenom Research Foundation, 3rd Floor, Narayana Health City, #258/A, Bommasandra, Hosur Road, Bangalore, Karnataka, 560 099, India; <sup>4</sup>Department of Pathology, Genentech Inc., 1 DNA WAY, South San Francisco, CA, 94080, USA; <sup>5</sup>Tea Research Association, North Bengal Regional R & D Centre, Nagrakata, Jalpaiguri, West Bengal, 735 225, India; <sup>6</sup>Pacific Biosciences, 1305 O'Brien Dr, Menlo Park, CA, 94025, USA; <sup>7</sup>SciGenom Labs Pvt Ltd, Plot no: 43A,SDF, 3<sup>rd</sup> floor, A Block, CSEZ, Kakkanad, Kochi, Kerala, 682 037, India.

\*Correspondence. S.S. sekar@gene.com; phone: 650-225-1000; fax: 650-225-1762

## Supplementary Figure legends

**Supplementary Figure S1:** Flowchart showing pipeline used for ORFs detection and annotation of the HytaNPV genome.

**Supplementary Figure S2:** Expression of HytaNPV ORFs at 24h and 72h using RNA-seq data.

**Supplementary Figure S3:** Comparison of HytaNPV against other baculoviruses used for biopesticides. Heatmap showing amino acid identity (%) resulting from blastp (evalue  $\leq 1$ ) of HytaNPV protein sequences against protein sequences of all complete baculovirus genomes.

**Supplementary Figure S4:** Alignment of repeat sequences from HytaNPV and that observed in BusuNPV.

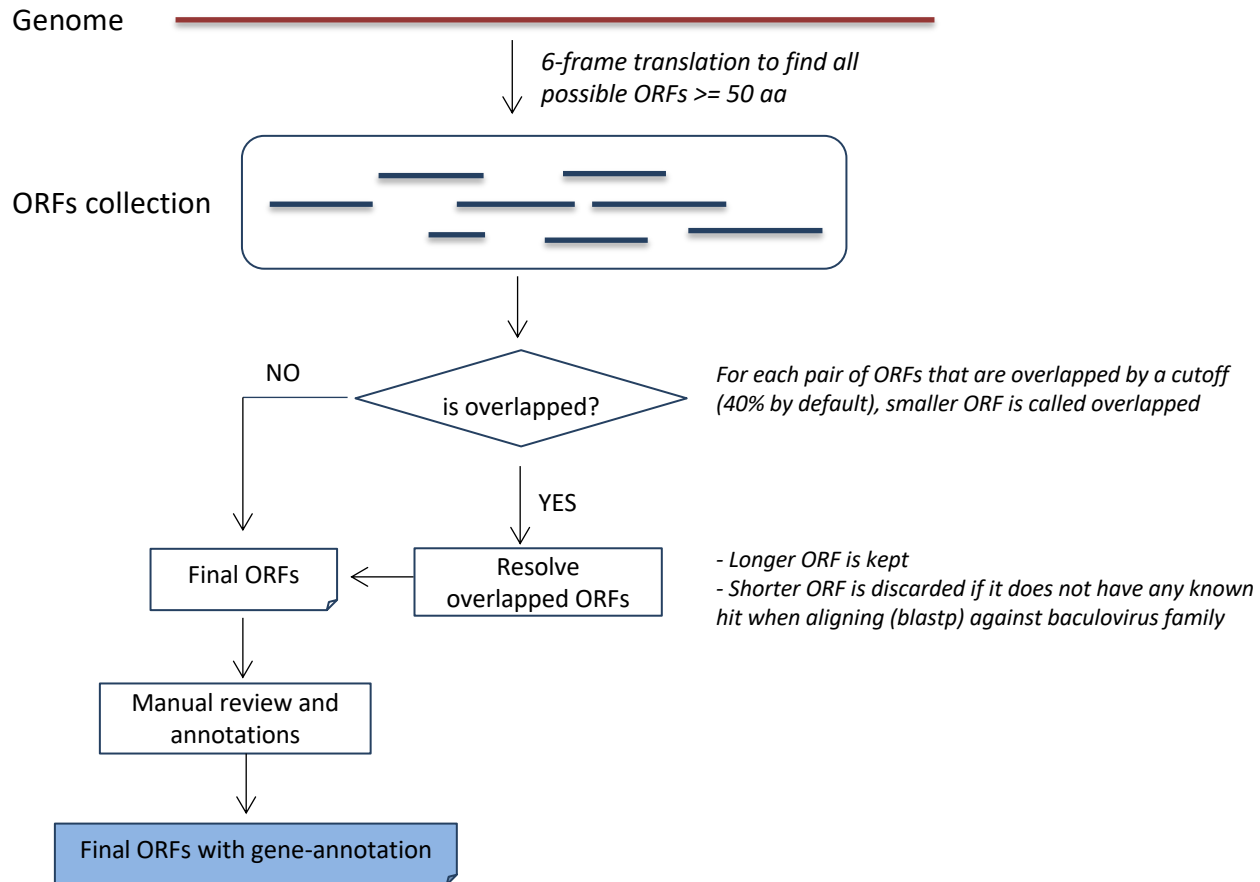
**Supplementary Figure S5:** Alignment of F and F-like proteins across baculoviruses. The multiple-sequence alignment was carried out using Clustal Omega<sup>1</sup> and visualized using Jalview software<sup>2</sup>. Cartoon of conserved regions (SP – signal peptide; FP – fusion peptide; S-S disulfide bridge; TM – transmembrane domain) of SeMNPV F protein shown at the bottom was adapted from Westenberg et al.<sup>3</sup>. Order of protein sequences appear on the left was determined by Clustal Omega's algorithm. A1 – Alphabaculovirus group I, B – Betabaculovirus; A2 – Alphabaculovirus group II. Amino acid conservation was highlighted in blue.

**Supplementary Figure S6:** Alignment of DNA sequences upstream of *polyhedrin* gene of AcMNPV, BusuNPV, SujuNPV and HytaNPV showing a conserved core promoter region.

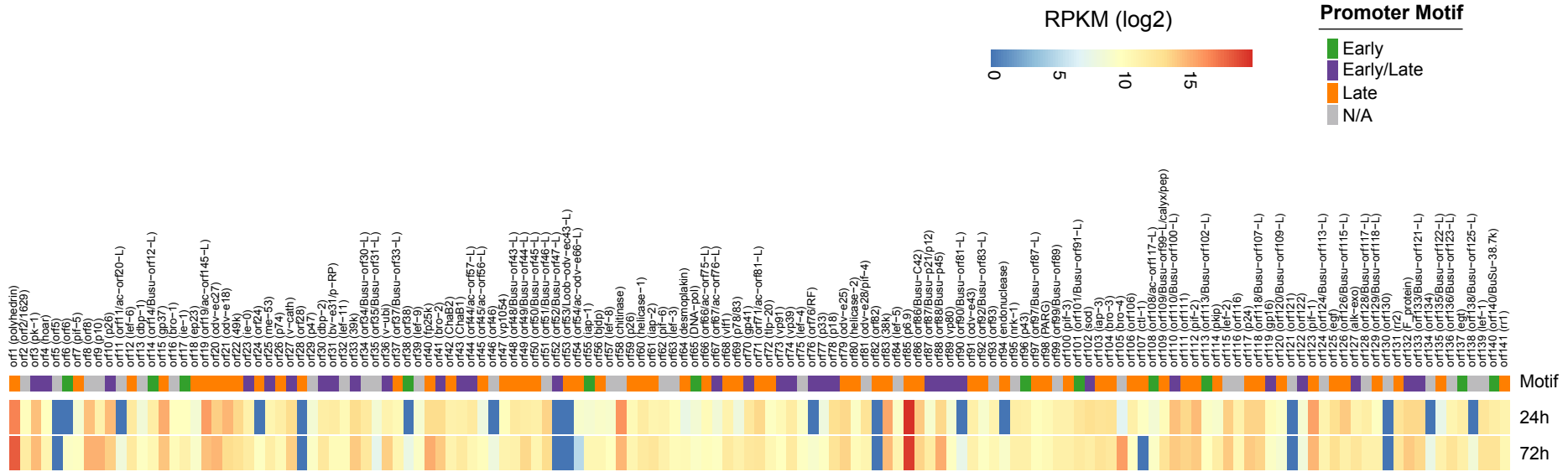
## Reference:

- 1 Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539, doi:10.1038/msb.2011.75 (2011).
- 2 Clamp, M., Cuff, J., Searle, S. M. & Barton, G. J. The Jalview Java alignment editor. *Bioinformatics* **20**, 426-427, doi:10.1093/bioinformatics/btg430 (2004).
- 3 Westenberg, M. *et al.* Functional analysis of the putative fusion domain of the baculovirus envelope fusion protein F. *J Virol* **78**, 6946-6954, doi:10.1128/JVI.78.13.6946-6954.2004 (2004).

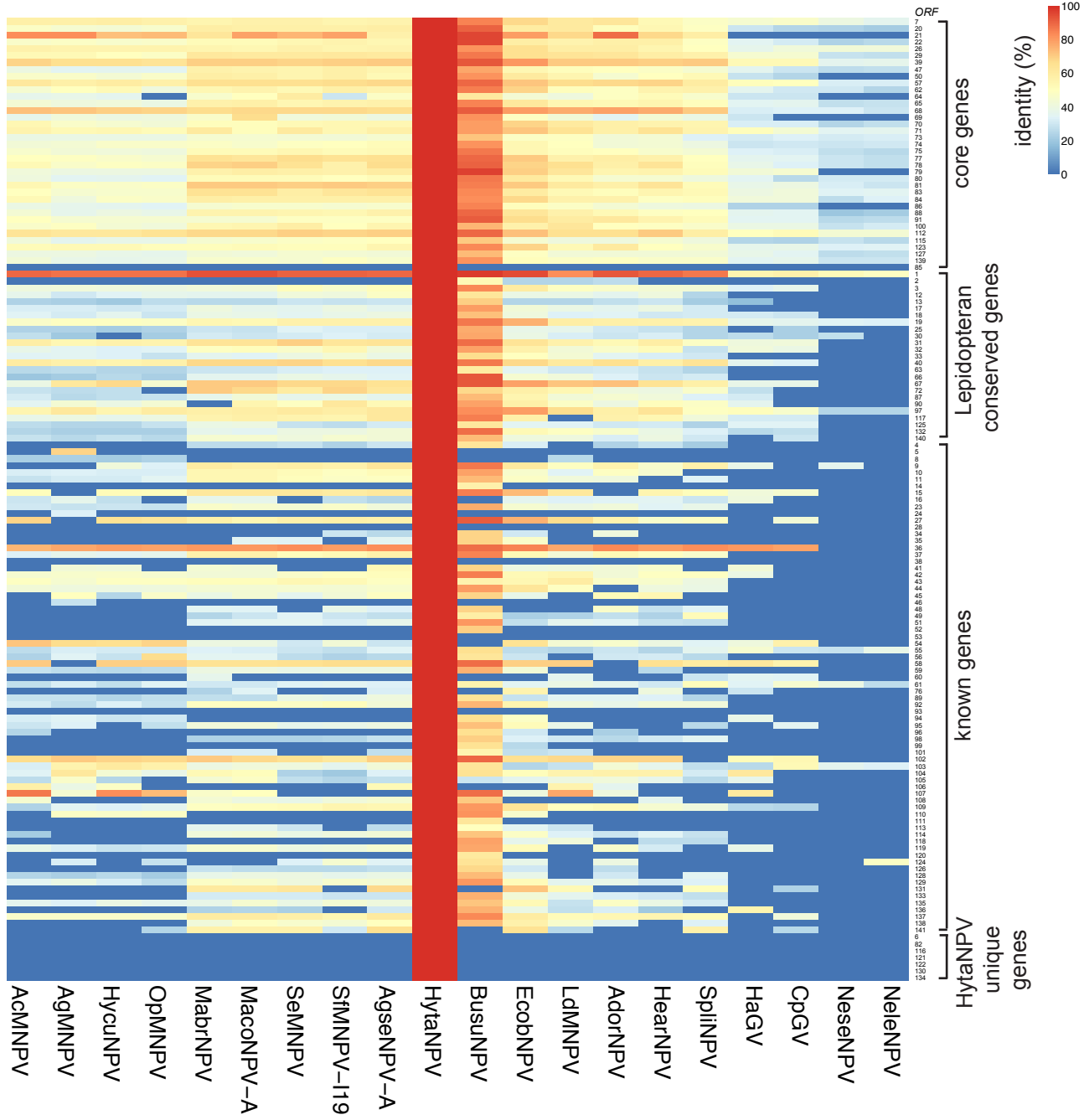
# Supplementary Figure S1



# Supplementary Figure S2



# Supplementary Figure S3



# Supplementary Figure S4

- repeat3-type1-46
- repeat3-type2-46
- repeat4 - type3-46
- repeat 5 -46
- repeat 6-46
- repeat 7-46
- busurepeat

```
TCGAAAATAACTTTTACTTTTAGATCTAACTGAAAGCCAAAATCTA
TCGAAAATAACTTTTACTTTTAGATCTAACTGAAAGCCAGAAATTC
TGACTTTTACTTTTGGATCTAACTAAAAGCAAGAAATCTATCGAATA
AGAAATCTATCGAAAATAACTTTTACTTTTAGATCTAACTAAAAGCC
AAAGCCAGAAATTTATCGAAAATGACTTTTACTTTTAGATTGAACTG
AAAGCCAGAAATTTATCGAAAATGACTTTTACTTTTAGATTGAACTA
CAGTATCAAAATATTTAGACT:TCTCGATTGTTGCTTTACTTTTAGATTTAACTGAAAG
```

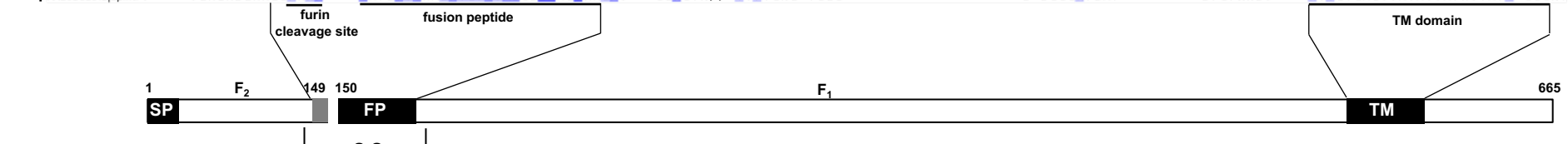
# Supplementary Figure S5

A1

B

A2

AGA16177.1 ThorMNPV	E-FAAYDDDD	SSSGGGGDI	AEQGHWSNLT	SDAQNLLR	NP-PKD	HFYPYK	IYINTASE	EK-I	QDSNNT	TIIVSII	IIVALVLF	GF-VIFLYFCV		
ABL75966.1 HavMNPV	E-FVYDYE	HLSHWSNMT	VSQAQALLO	NP-PKD	HFYRPSK	IYNTSATS	DE	KDNNIN	IVIMI	IAVAALL	LLCGL	MIIFLFCI		
AC057205.1 BomMalNPV	D-FVYDYE	PSHWSNMT	VSQAQALLO	NP-PKD	HFHQSK	IYNTSATS	DSDS	QDSSIT	TVVVI	IAVAAL	FLFCSL	SLFLFCI		
AC63697.1 BmMNPV	D-FVYDYE	PSHWSNMT	VSQAQALLO	NP-PKD	HFHQSK	IYNTSATS	DSDS	QDSSNIT	TVVVI	IAVAAL	FLFCSL	LLFLFCI		
AAM28022.1 RoMNPV	E-FVYDYE	QPSHWSNMT	VSQAQTLNR	NP-PKD	HFHQSK	IYNTSAAAP	DSDS	QDDNNT	IIVIV	IAVAIL	LLFCGL	LFLYCCI		
AA66653.1 AcMNPV	E-FVYDYE	QPSHWSNMT	VSQAQALLR	NP-PKD	HFHQSK	IYNTSAAAP	DSDS	QDSSNTT	VVIA	IAVAAM	LLFCGL	LLFLFCI		
ABE68408.1 PxyMNPV	E-FVYDYE	QPSHWSNMT	VSQAQALLR	NP-PKD	HFHQSK	IYNTSAAAP	DSDS	QDSSNTT	VVIA	IAVAAM	LLFCGL	LLFLFCI		
AAK85583.1 EppMNPV	D-FNYIDNR	QENVYDNT	HNVHATQ	LNISEAR	ILNLTDL	TKD	PFSPHN	DLPTO	P	IYVTVI	SALCVVMFY	CMY		
BAE72419.1 HyuMNPV	DFVYDTR	PQNIIDVNT	KNKPRWS	LLSAGVRELLT	AD-ASD	VFYSQR	INVPD	TKPA	QNNANV	VLI	SVLVSV	GFSAFCV	IYAGYCLL	
AGR57168.1 ChroMNPV	N-1EYDAP	QTD	VDAHNP	PHWSK	VSAAVKA	VLA	AP-SRK	SYFSPRR	NDVGA	H	EPPT	QNKVNVV	VLV	
AKR14213.1 DapuMNPV	E-VDYDTR	QDDAA	FADAHNP	PHWSV	VSAAVKE	LLA	VAPPRD	TFAPRR	SDAAA	P	DPH	EK	PT	
AC59020.2 OpMNPV	E-VDYDTR	QDDAA	FADAHNP	PHWSV	VSAAVKE	LLA	VAPPRD	TFAPRR	SDAAA	P	DPH	EK	PT	
AHD25615.1 ChmuMNPV	E-FDYDTR	QDDTA	FADAHNP	PHWST	VGAAD	IKALLA	APPPRH	RFFSPRR	NDVNA	P	ELPV	QNEHNI	VLV	
AA29812.3 CfMNPV	E-FDYDTR	QDDTA	FADAHNP	PHWST	VGAAD	IKALLA	APPPRH	RFFSPRR	NDVNA	P	ELPV	PNEHNI	VLV	
AGR57018.1 ChocNPV	E-FDYDTR	QDDTA	FADAHNP	PHWST	VGAAD	IKALLA	APPPRH	RFFSPRR	NDVNA	P	ELPV	PNEHNI	VLV	
ABF50356.1 AnpeNPV	F-FDYDTR	QDDAD	DEASHAH	WSDVNA	AAARALLR	GA-AAN	LFFSPRR	DV-OA	P	EPYV	DTNTGAV	L		
AFY62925.1 PhcyNPV	F-FDYDTR	QDDAD	DEASHAH	WSDVNA	AAARALLR	GA-AAN	LFFSPRR	DV-OA	P	EPYV	DTNTGAV	L		
AJD09285.1 CoveNPV	F-FDYDTR	QDDADY	ADTNP	PHWSE	VSMVDR	ALMR	GG-STN	NFFALRN	NV-HS	P	QIVD	KNVDLTY	S	
ABI13806.1 AgMNPV	F-FDYDTR	QDDADY	ADTNP	PHWSE	VSMVDR	ALMR	GG-STN	NFFALRN	NV-HS	P	QIVD	KNVDLTY	S	
AA091645.1 CIDEFNPV	F-FDYDTR	QDDADY	ADTNP	PHWSE	VSMVDR	ALMR	GG-STN	NFFALRN	NV-HS	P	QIVD	KNVDLTY	S	
ANF29764.1 CapoNPV	I-YDDNDG	LDVKKANG	INGDGG	IDGNNS	KWVLT	ENEAR	TILORNET	SKHNGVDG	NRFF	TFPOHNY	AVLHHP	FNTKNAVD	ININ	
ANF29765.1 JekMNPV	I-YDDNDG	SRELKTR	MTAA	AAAAA	TAAADR	LRHWS	ELSN	ENALM	RR	STSEK	RFAPSK	IYTL	NV	
AKN81073.1 LooMNPV	Y-EDDY	DGSGDD	KFLP	ARDQ	HWS	ELN	TARQL	MLP	NS	ISA	LVTR	GYVMNVV	I	
AKC91684.1 IlyspNPV	---	AAAAGH	KRNV	FGGAF	NFVGR	VKY	LF	GVMDD	DA	ELLYK	LA	QSNTH	H	
AKR17408.1 MolaGV	---	PVRNG	SKRNI	FGGAF	NFVGR	VKY	LF	GVMDD	DA	ELLYK	LA	QSNTH	H	
ACH69376.1 PsunGV	---	PLKNS	PKRNI	FGGAF	NFVGR	VKY	LF	GVMDD	DA	ELLYK	LA	QSNTH	H	
AOW41364.1 TnGV	---	PLKNS	PKRNI	FGGAF	NFVGR	VKY	LF	GVMDD	DA	ELLYK	LA	QSNTH	H	
ABY47717.1 HaGV	---	PVKNG	AKRNI	FGGAF	NFVGR	VKY	LF	GVMDD	DA	ELLYK	LA	QSNTH	H	
AAF05141.1 XcGV	---	PVKNG	AKRNI	FGGAF	NFVGR	VKY	LF	GVMDD	DA	ELLYK	LA	QSNTH	H	
AKN80710.1 DisaGV	---	EEEST	RKR	LRG	---	---	---	---	---	---	---	---	---	
AB051967.1 SpiIGV	---	VRV	KRNL	FNGA	FVGR	FDKY	LF	GVMDD	DA	ELLYK	LA	QSNTH	H	
AA085660.1 AadorGV	---	VDV	NL	GM	LN	FVGR	TKY	TI	GM	DD	DA	ELLYK	LA	
AA627324.1 PlyxGV	---	KRQR	ELL	GA	FVGR	FNY	VY	GM	DD	DA	ELLYK	LA	QSNTH	H
AE41440.1 EpapGV	---	R	SKR	I	FGGT	NFVGR	VKY	LF	GVMDD	DA	ELLYK	LA	QSNTH	H
AMF83797.1 CnmeGV	---	KRKR	Q	SWF	SGT	NFVGR	FKNM	I	GM	DD	DA	ELLYK	LA	
AE800318.1 CianGV	---	RMKR	LN	FNGA	FVGR	VKY	LF	GVMDD	DA	ELLYK	LA	QSNTH	H	
AG020290.1 ClasGV/B	---	R	KRNL	FNGA	FVGR	VKY	LF	GVMDD	DA	ELLYK	LA	QSNTH	H	
AA852713.1 AgseGV	---	R	KRS	I	FGGAF	NFVGR	VKY	LF	GVMDD	DA	ELLYK	LA	QSNTH	H
AKS25366.1 ClasGV/A	---	GRV	KRNL	FNGA	FVGR	VKY	LF	GVMDD	DA	ELLYK	LA	QSNTH	H	
AKS92028.1 EreiGV	---	NRV	KRNL	FNGA	FVGR	VKY	LF	GVMDD	DA	ELLYK	LA	QSNTH	H	
AP013911.1 PiGV	---	R	RKR	I	FGGAF	NFVGR	VKY	LF	GVMDD	DA	ELLYK	LA	QSNTH	H
AA021625.1 CrieGV	---	TRR	KR	GL	FGGAF	NFVGR	VKY	LF	GVMDD	DA	ELLYK	LA	QSNTH	H
AAK70691.1 CpGV	---	RHR	KR	GL	FGGAF	NFVGR	VKY	LF	GVMDD	DA	ELLYK	LA	QSNTH	H
AAK70225.1 PhocGV	---	K	RL	FGGAF	NFVGR	VKY	LF	GVMDD	DA	ELLYK	LA	QSNTH	H	
ABC61157.1 ChocNPV	---	S	RT	K	RG	FGGAF	NFVGR	VKY	LF	GVMDD	DA	ELLYK	LA	
AC63512.1 PiraGV	---	NRRT	K	RG	FGGAF	NFVGR	VKY	LF	GVMDD	DA	ELLYK	LA	QSNTH	H
AKC91636.1 IafMNPV	NGEAAA	SKPA	KRRKR	---	---	---	---	---	---	---	---	---	---	
BAC67369.1 AdhoNPV	---	KVRNKR	---	---	---	---	---	---	---	---	---	---	---	
ACF05406.1 AadorNPV	---	KRNRKR	---	---	---	---	---	---	---	---	---	---	---	
ACQ70316.1 LdMNPV	---	LAPRRKR	---	---	---	---	---	---	---	---	---	---	---	
ADD73835.1 LxyxMNPV	---	LATRKR	---	---	---	---	---	---	---	---	---	---	---	
AA95955.1 CibiNPV	---	GGDV	HRRAKR	---	---	---	---	---	---	---	---	---	---	
AKN80684.1 PelusMNPV	GGGVA	I	GGDK	QRRKR	---	---	---	---	---	---	---	---	---	
ABY65853.1 OrlenNPV	QDVAQR	A	TRRKR	---	---	---	---	---	---	---	---	---	---	
AC053584.1 EupsNPV	L	TG	KAVR	LVK	QROKR	---	---	---	---	---	---	---	---	
AU41354.1 SuijNPV	TNDR	I	VRHKR	---	---	---	---	---	---	---	---	---	---	
orf.32 HytaNPV	TNKNL	K	NRKR	---	---	---	---	---	---	---	---	---	---	
AHH82709.1 BusiNPV	TNKNL	K	NRKR	---	---	---	---	---	---	---	---	---	---	
ABI35801.1 EcoBpNPV	D	H	ES	V	M	S	K	R	R	R	R	R	R	
AD844371.1 ApoNPV	S	SP	K	I	L	N	R	R	K	R	---	---	---	
AGR56881.1 HespNPV	S	R	E	S	A	E	R	G	H	L	R	R	R	
AAK96377.1 HearNPV	V	T	D	K	P	T	F	S	---	---	---	---	---	
AKR17392.1 UrrpNPV	S	T	K	A	P	S	F	N	A	K	R	R	R	
AF164959.1 ImabNPV	P	T	---	---	---	---	---	---	---	---	---	---	---	
AAM091117.1 MacoNPV/A	P	P	---	---	---	---	---	---	---	---	---	---	---	
AY84081.1 ChroMNPV	S	R	K	R	---	---	---	---	---	---	---	---	---	
AJD08030.1 ChinSNPV	---	N	R	S	K	R	---	---	---	---	---	---	---	
AZ67513.1 TnNPV	---	S	R	S	K	R	---	---	---	---	---	---	---	
AE47740.1 PespNPV	---	L	K	T	H	R	R	K	---	---	---	---	---	
AZ382174.1 AgseNPV-A	DE	---	P	Y	P	S	R	R	R	---	---	---	---	
AC128714.1 AgipMNPV	E	---	D	D	S	K	R	K	---	---	---	---	---	
AIZ48567.1 AgseNPV-B	E	---	N	D	D	S	K	R	K	---	---	---	---	
AAF33539.1 SemNPV	S	---	P	T	K	R	---	---	---	---	---	---	---	
ABM45723.1 SImNPV-19	E	---	K	T	N	T	---	---	---	---	---	---	---	
AAR28930.1 LeseNPV	---	D	A	P	P	R	R	R	---	---	---	---	---	
AA01818.1 SpilNPV	V	E	R	L	D	E	R	L	R	S	R	R	R	





# Supplementary Figure S6

Reference sequence (1): AcMNPV  
 Identities normalised by aligned length.  
 Colored by: identity + property

		1 [	.	.	.	:	.	.
1 AcMNPV	100.0%	ac	tattgtctgcgagcagttg	tttgt--	tgtttaa	-aataacagccattg	taatgagacgc--	acaaactaatatcac--
2 BusuNPV	48.6%	-	ctttaa-atgctgctgttg	gcacat--	cgatctgtgttacat---	gttgaaacg-gatg-agat--	ctcaatattttga	
3 HytaNPV	50.0%	-	attgta-acgcgtactgtcgt-	cg--	cacactgacgtttaa	---attaacattacatg-	atattatgcattatcataa	
4 SujuNPV	48.3%	---	tta-atgtaaat--	ttg	tacgtgagaaaataataaaaac	ctttgatactactgtat	gttttttgat	ttt-tta
consensus/100%		...	sTu.ssGsssss..TsGs.suT..s	sssssu.suTssss...	sTTussAss.sssG...	ss..sssusTsTs.s...		
consensus/90%		...	sTu.ssGsssss..TsGs.suT..s	sssssu.suTssss...	sTTussAss.sssG...	ss..sssusTsTs.s...		
consensus/80%		...	sTu.ssGsssss..TsGs.suT..s	sssssu.suTssss...	sTTussAss.sssG...	ss..sssusTsTs.s...		
consensus/70%			sTSTTA ATGcuuucSGTTGtsCGT	ssssTusAuTAAAA...	uTTGAsAssAsATG.As	AssssssAATATssTsA		

		81	.	.	.	:	.	.
1 AcMNPV	100.0%	aaactggaaatg	tc-tatcaata-tatag	ttgct-gat	atcatggag--	ataat-taaaatgata	accatctcgcaa--	
2 BusuNPV	48.6%	acacctg--	acggctttacagca-tg	taattgca-tat	acgatgtttata	aaattatcaaaat	ttgttcatt-tacaatc	
3 HytaNPV	50.0%	atactgg--	acggctt	atcagtg--	cgctcgctcta	acgatg	ttttgtaaatt-ttaaa	ttttgttcatt-tacaac
4 SujuNPV	48.3%	aaactat--	atg	tat-ccc	atgacaataatcgat-	tat	atgattaattgtaaaat-ttaaa	att-cttcagt-cacaacg
consensus/100%		AsACsss..AsGss..	ssCAssu...susTsGss.s	ATAssATssss...	AsAsT.Ts	AAAsss.sssCAss.suCAA..		
consensus/90%		AsACsss..AsGss..	ssCAssu...susTsGss.s	ATAssATssss...	AsAsT.Ts	AAAsss.sssCAss.suCAA..		
consensus/80%		AsACsss..AsGss..	ssCAssu...susTsGss.s	ATAssATssss...	AsAsT.Ts	AAAsss.sssCAss.suCAA..		
consensus/70%		AsACTuG AsGscTTssCAUsa	suTAuTsGCT	TATAsGATGssTtuT	AAAsT Ts	AAAsTTtuTTCAAT	sACAAss	

		161	.	.	.	2	.	217
1 AcMNPV	100.0%	ataaataagta	ttttactg	ttttcgtaacag	ttttgt	aataa-aaaac	tataaat	
2 BusuNPV	48.6%	ttcaataagta	tttttttctc	tattgtaaaaca-tt	gtgaaaaat	caaatacaacata		
3 HytaNPV	50.0%	ttcaataagta	tttttttctc	tattgtaaaaca-tt	gtgaaaaat	caaatacaacata		
4 SujuNPV	48.3%	ttcaataagta	ttttttgttc	tattgtaaaaca-tt	gtgaaaaat	caaatac-acata		
consensus/100%		sTsAA	AAGTATTTTTsTsTs	GTAAsAss.TTG	TuAsAA.ss	AAAsss.ssAss		
consensus/90%		sTsAA	AAGTATTTTTsTsTs	GTAAsAss.TTG	TuAsAA.ss	AAAsss.ssAss		
consensus/80%		sTsAA	AAGTATTTTTsTsTs	GTAAsAss.TTG	TuAsAA.ss	AAAsss.ssAss		
consensus/70%		TTCAAT	AAGTATTTTTsTsC	TATTGTA	AAAAACA TTG	TGAAAAATCAAATAs	AACATA	

