

Supplementary Material

Supplementary table 1. Predicted ORFs and their functional annotation in the genome of *Pseudomonas* phage Eir4. Putative Shine-Dalgarno sequences 20 bases upstream of the predicted start codon of each ORF that are complementary to the 16S rRNA tail of *Pseudomonas syringae* pv. tomato (host) are underlined and highlighted in bold. Color coding for ORF product functional groups is as follows: yellow - proteins involved in virion morphogenesis; red—DNA-dependent RNA polymerase; blue —DNA replication, modification and repair; green — lysis; purple — additional functions; gray — hypothetical proteins of unknown function.

ORF	Strand	Start position (bp)	End position (bp)	Product size (aa)	Shine-Dalgarno sequence		Start codon	Protein accession	Predicted function
					20 bp upstream of start codon (G-U wobble base pairs allowed)	DeltaG (kcal/mol; as calculated by free_align.pl -o "20bp" AUUCCUCCACUAG)			
#1	+	117	278	53	GGTTCCTTC <u>GAGGG</u> CATCTC	-4,046925	ATG	UGL61051.1	hypothetical protein
#2	+	434	559	41	<u>ATGGTT</u> GACAATCCTCTGAG	-2,18726	ATG	UGL61052.1	hypothetical protein
#3	+	809	934	41	GGCAACCATT <u>TGAGG</u> ACTCAC	-5,87344	TTG	UGL61053.1	hypothetical protein
#4	+	931	1185	84	TCACTG <u>AAGGA</u> TGCAACATC	-4,53508	ATG	UGL61054.1	hypothetical protein
#5	+	1185	1382	65	GCTGAC <u>AAGGAG</u> CCTGAGTG	-6,610015	ATG	UGL61055.1	hypothetical protein
#6	+	1387	1698	103	TCACCAG <u>AAGGG</u> CTAACATC	-3,71267	ATG	UGL61056.1	hypothetical protein
#7	+	1788	2267	159	CAACAAACGA <u>AGGT</u> AGCACC	-3,49838	ATG	UGL61057.1	hypothetical protein
#8	+	2298	2840	180	CGCTGTGAT <u>AGGTG</u> TTTTCA	-5,595345	ATG	UGL61058.1	hypothetical protein
#9	+	2837	3541	234	ACCTTTC <u>AGGGG</u> ATGACAGC	-4,90869	ATG	UGL61059.1	putative hydroxylase
#10	+	3538	3801	87	GTGCATT <u>GGAGT</u> GAAATCAA	-3,60793	GTG	UGL61060.1	hypothetical protein
#11	+	3804	4403	199	TACC <u>GGGGAGG</u> CTTCTAACC	-9,478722	ATG	UGL61061.1	hypothetical protein
#12	+	4543	7200	885	AACCCATT <u>CGAG</u> ACGACTTA	-0,359835	ATG	UGL61062.1	DNA-directed RNA polymerase
#13	+	7214	7351	45	TGCAT <u>TAAGGA</u> AACCAATCAA	-5,867005	ATG	UGL61063.1	hypothetical protein
#14	+	7348	7617	89	CGTC <u>AGGAGT</u> GGAACGAGGA	-5,682865	ATG	UGL61064.1	hypothetical protein
#15	+	7617	8000	127	TTCCTG <u>AAGGAG</u> ACCAATTA	-6,610015	ATG	UGL61065.1	hypothetical protein
#16	+	8012	9079	355	GCCAAG <u>TAAGGAG</u> CAACCCA	-7,94194	ATG	UGL61066.1	ATP-dependent DNA ligase
#17	+	9273	9530	85	CCCTCA <u>AAGGAG</u> ACTCAACG	-6,610015	ATG	UGL61067.1	hypothetical protein
#18	+	9527	10171	214	ATCGT <u>GAAGGAGGTG</u> TTCGC	-14,20565	ATG	UGL61068.1	kinase

Supplementary Material

#19	+	10168	10338	56	CGCAACCAAGGAGTAATCGA	-6,610015	ATG	UGL61069.1	bacterial RNA polymerase inhibitor
#20	+	10335	10697	120	CTGCGCCCAAGGTGACTCA	-10,315595	ATG	UGL61070.1	hypothetical protein
#21	+	10749	11450	233	ATCTGCAAGGAGAAACACA	-6,610015	ATG	UGL61071.1	single-stranded DNA-binding protein
#22	+	11450	11893	147	TCCGGTCACGGTGACTTCTG	-5,880535	ATG	UGL61072.1	endonuclease I
#23	+	11895	12335	146	ACCAAAGAAGGAGGTTGAC	-12,108685	ATG	UGL61073.1	lysozyme
#24	+	12411	12953	180	GTTGACATAAGGAGGTTCT	-13,44061	ATG	UGL61074.1	putative nucleotidyltransferase
#25	+	12940	14631	563	GAATGAAGTGGAGGTGGCTG	-12,74128	ATG	UGL61075.1	DNA primase/helicase
#26	+	14645	14848	67	CTACTAAGGAGAACCATCAC	-7,94194	ATG	UGL61076.1	hypothetical protein
#27	+	14914	15375	153	CTCAATAGGAGAACCACACC	-5,682865	ATG	UGL61077.1	hypothetical protein
#28	+	15386	17530	714	CCAACGATAGGAGGACGCTG	-8,93096	ATG	UGL61078.1	DNA polymerase
#29	+	17543	17926	127	GTCACATAAGGAGGCCCTA	-8,93096	ATG	UGL61079.1	hypothetical protein
#30	+	17919	18128	69	GATCGCCCGAGGTGGCAGCG	-9,493185	ATG	UGL61080.1	hypothetical protein
#31	+	18138	19088	316	ACGCCATCTAAGGAGGGCCT	-11,190035	ATG	UGL61081.1	exonuclease
#32	+	19155	19397	80	TATTCGTAAAGGAGGGCTA	-9,85811	ATG	UGL61082.1	DUF2717 domain-containing protein
#33	+	19400	19684	94	CTGCAAAGGGAGGAATGACC	-9,521905	ATG	UGL61083.1	DUF5476 domain-containing protein
#34	+	19681	20124	147	GCCCTGAAGAAGGTCACCAA	-3,49838	GTG	UGL61084.1	N-acetyltransferase
#35	+	20096	20401	101	GAAACTGAAGGAGAACCCCA	-6,610015	ATG	UGL61085.1	tail assembly protein
#36	+	20415	22040	541	CATTTAATTGGAGGTGACCA	-13,56369	GTG	UGL61086.1	portal protein
#37	+	22110	22985	291	ATTCCCTCAAGGAGACCTCA	-6,610015	ATG	UGL61087.1	capsid assembly protein
#38	+	23088	24131	347	CATGCAAGGAGAACTACAC	-6,610015	ATG	UGL61088.1	major capsid protein
#39	+	24156	24779	207	ACCCAACCCCTTGAGGCCCT	-3,60793	TTG	UGL61089.1	tail tubular protein A
#40	+	24789	27209	806	CAAACCGATAAGGAGGGCCT	-11,190035	ATG	UGL61090.1	tail tubular protein B
#41	+	27284	27718	144	CTACGGGAGGGTTTCAACTT	-8,979195	ATG	UGL61091.1	acetyltransferase
#42	+	27729	28310	193	CTGGCTGTAGGAGGTACACC	-11,181535	ATG	UGL61092.1	internal virion protein B
#43	+	28310	30526	738	CCACAAGGAGGCAAGACGTA	-9,85811	ATG	UGL61093.1	internal virion protein C
#44	+	30539	34729	1396	CTCTCTAATAGGAGGTCTAC	-11,181535	ATG	UGL61094.1	internal virion protein D
#45	+	34802	36550	582	TTCAACCACAAGGAGACTCT	-6,610015	ATG	UGL61095.1	tail fiber protein
#46	+	36587	36946	119	GGCCTCTCTAAGGAGGACTT	-11,190035	ATG	UGL61096.1	hypothetical protein

#47	+	36946	37161	71	CAGATC <u>AAGG</u> CGATGAGCTG	-2,174955	ATG	UGL61097.1	holin
#48	+	37158	37415	85	AAGAACA <u>AAGGAG</u> ACCCACC	-6,610015	ATG	UGL61098.1	terminase small subunit
#49	+	37418	37873	151	GACGGCGT <u>GAGG</u> CACTGACA	-3,60793	GTG	UGL61099.1	i-spanin
#50	+	37569	37823	84	ACGACAACT <u>GGAGGT</u> TGAAC	-9,1066	GTG	UGL61100.1	o-spanin
#51	+	37883	39634	583	CGCACAAAT <u>TAAGGAGG</u> CCCA	-11,190035	ATG	UGL61101.1	terminase large subunit
#52	+	39813	39986	57	CATCA <u>AAGGAGG</u> ACAAGACC	-9,85811	ATG	UGL61102.1	hypothetical protein

Supplementary table 2. Predicted ORFs and their functional annotation in the genome of *Pseudomonas* phage Eisa9. Putative Shine-Dalgarno sequences 20 bases upstream of the predicted start codon of each ORF that are complementary to the 16S rRNA tail of *Pseudomonas syringae* pv. tomato (host) are underlined and highlighted in bold. Color coding for ORF product functional groups is as follows: yellow - proteins involved in virion morphogenesis; red—DNA-dependent RNA polymerase; blue —DNA replication, modification and repair; green — lysis; purple — additional functions; gray — hypothetical proteins of unknown function.

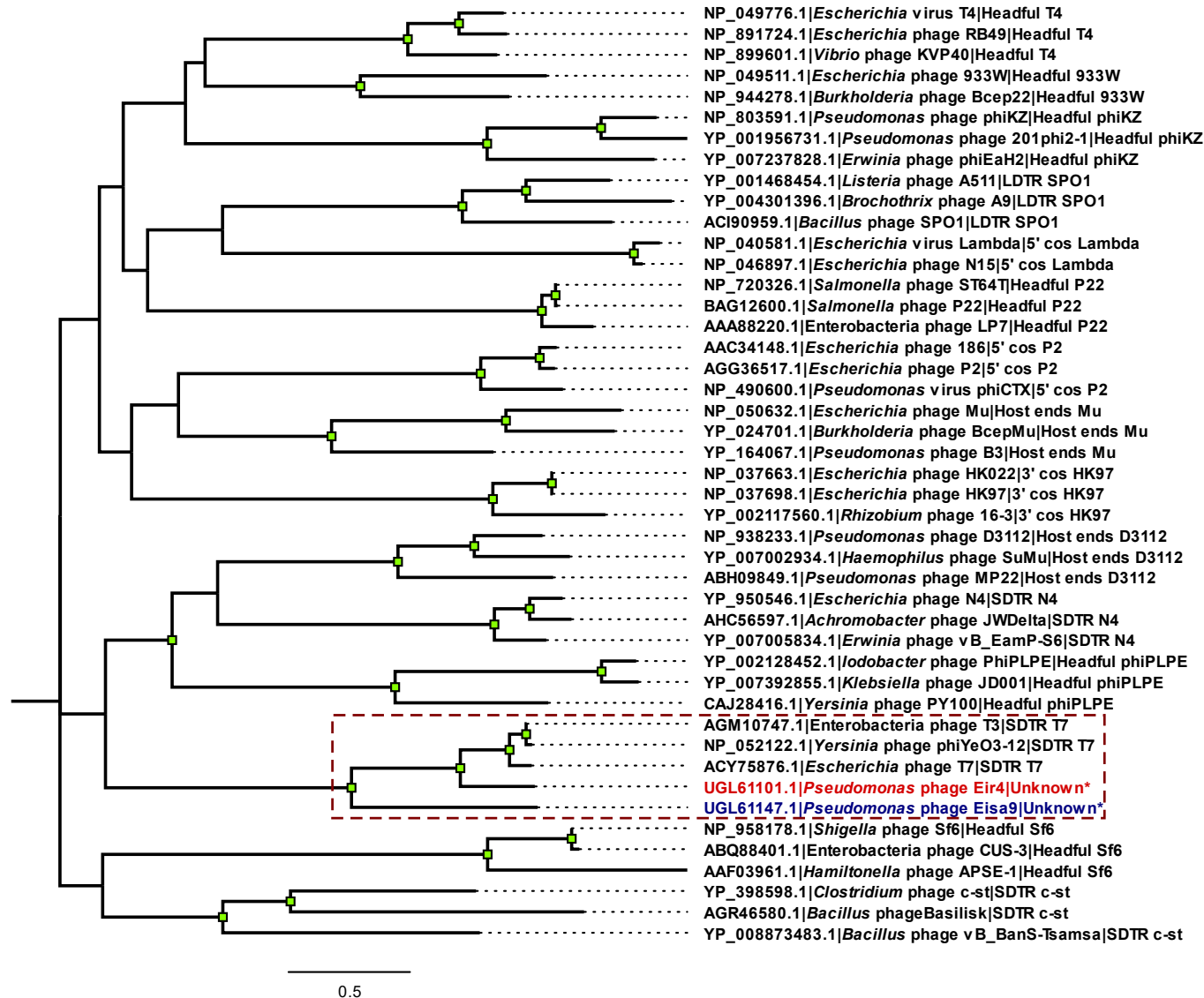
ORF	Strand	Start position (bp)	End position (bp)	Product size (aa)	Shine-Dalgarno sequence		Start codon	Protein accession	Predicted function
					20 bp upstream of start codon (G-U wobble base pairs allowed)	DeltaG (kcal/mol; as calculated by free_align.pl -o "20bp" AUUCCUCCACUAG)			
#1	+	21	236	71	AT <u>AGG</u> CCAAGTGTAGTAAG	-1,247805	TTG	UGL61103.1	hypothetical protein
#2	+	353	619	88	GCAACCACT <u>GAG</u> TAACGACA	-0,381865	ATG	UGL61104.1	hypothetical protein
#3	+	647	1147	166	AACTGACT <u>TGAGGG</u> TTTCATC	-5,05103	ATG	UGL61105.1	hypothetical protein
#4	+	1208	1480	90	TTTGCAT <u>AGGAG</u> ACAGGCGC	-5,682865	ATG	UGL61106.1	hypothetical protein
#5	+	1482	1709	75	CGCCAATCGCC <u>GGGG</u> CTGAC	-2,833755	ATG	UGL61107.1	hypothetical protein
#6	+	1709	1876	55	ACTAACTAC <u>GAGGTGGT</u> GTA	-10,859585	ATG	UGL61108.1	hypothetical protein
#7	+	1869	2192	107	CCAAGTGCC <u>GGGTG</u> CTGACT	-5,64358	ATG	UGL61109.1	hypothetical protein
#8	+	2449	2946	165	GCAACAACATC <u>AGGTGA</u> ACT	-7,95547	ATG	UGL61110.1	putative acetyltransferase
#9	+	2937	3359	140	GTATTCAT <u>GGAG</u> CTGCCAGC	-3,60793	ATG	UGL61111.1	hypothetical protein
#10	+	3343	3537	64	GATAGCGGC <u>GGA</u> CCAAGAAC	-1,532995	ATG	UGL61112.1	hypothetical protein
#11	+	3534	3794	86	ACCTTC <u>AAGGAGT</u> ACAGGCC	-6,610015	ATG	UGL61113.1	hypothetical protein
#12	+	3791	4669	292	CTGATGAT <u>GGAGT</u> GTGACGA	-3,60793	ATG	UGL61114.1	hypothetical protein
#13	+	4669	4977	102	CGTCAAGT <u>GGAG</u> TTCGTGTA	-3,60793	ATG	UGL61115.1	hypothetical protein
#14	+	4982	5242	86	CACTCT <u>GGAGGG</u> GCTAAGCCC	-6,856025	ATG	UGL61116.1	hypothetical protein
#15	+	5325	5807	160	CCAAACGA <u>AAGGA</u> TACACC	-4,53508	ATG	UGL61117.1	hypothetical protein
#16	+	5807	6514	235	GCCCT <u>GATC</u> GGCAGCCTGTA	-3,840985	ATG	UGL61118.1	hypothetical protein
#17	+	6511	7314	267	CTGGTGTCT <u>GGAGT</u> AAGCA	-3,60793	TTG	UGL61119.1	putative DNA primase
#18	+	7480	7668	62	TCCGCAC <u>AGGAG</u> ACGCGAGC	-5,682865	ATG	UGL61120.1	hypothetical protein
#19	+	7658	8944	428	TAACTGGT <u>GGAGGT</u> ACACG	-9,1066	ATG	UGL61121.1	DNA helicase
#20	+	8931	9161	76	GGATGCAAC <u>GGGAGG</u> ACGTT	-8,979195	ATG	UGL61122.1	hypothetical protein
#21	+	9158	11494	778	CTGTGCAAGGCGT <u>GGTG</u> CGC	-3,52041	ATG	UGL61123.1	DNA polymerase
#22	+	11503	12390	295	CAATACG <u>AGTGAG</u> AAGAAGAC	-2,63244	ATG	UGL61124.1	hypothetical protein

#23	+	12390	13400	336	GCGCTGGCC <u>GGT</u> ATCCTCTG	-1,423445	ATG	UGL61125.1	exonuclease
#24	+	13402	13743	113	TCGCACCGATCA <u>GGG</u> TAAA	-3,546615	ATG	UGL61126.1	putative recombination endonuclease
#25	+	13743	14717	324	AAAGCAA <u>AGG</u> TATCGAAATA	-3,49838	ATG	UGL61127.1	hypothetical protein
#26	+	14717	15274	185	CGCTGGGTC <u>GGGAG</u> CCTCTA	-5,7311	ATG	UGL61128.1	DUF5664 domain-containing protein
#27	+	15261	15446	61	TCTCCTGT <u>TGAGG</u> ATCAAAC	-5,87344	ATG	UGL61129.1	hypothetical protein
#28	+	15446	16348	300	ATCTCT <u>GGAGGTGA</u> AATCTG	-13,56369	ATG	UGL61130.1	DNA ligase
#29	+	16349	18772	807	AGACCGAGCCG <u>AGTGAG</u> TAA	-2,63244	ATG	UGL61131.1	DNA-dependent RNA polymerase
#30	+	18830	19132	100	ACTTGCC <u>GGAGG</u> ATCGCTTC	-6,856025	ATG	UGL61132.1	hypothetical protein
#31	+	19119	19601	160	GCGGACAGT <u>GGAGG</u> CCGAAC	-6,856025	TTG	UGL61133.1	hypothetical protein
#32	+	19603	20079	158	CGTCAAAC <u>GGAG</u> TATCTGAT	-3,60793	ATG	UGL61134.1	hypothetical protein
#33	+	20076	21590	504	AGCG <u>GATCGAG</u> CGTCCGTCT	-1,74402	GTG	UGL61135.1	portal protein
#34	+	21587	22372	261	GTACCC <u>GGAG</u> ATACTGCACA	-3,60793	ATG	UGL61136.1	scaffolding protein
#35	+	22449	23456	335	CACCACAT <u>TAAGG</u> ATTCCACA	-5,867005	ATG	UGL61137.1	major capsid protein
#36	+	23531	24133	200	TTGCGTTT <u>AGGAG</u> ATAGCCT	-5,682865	ATG	UGL61138.1	tail tubular protein A
#37	+	24130	26661	843	CCGAC <u>AGGAC</u> GCGCATTGAG	-3,60793	ATG	UGL61139.1	tail tubular protein B
#38	+	26665	27414	249	TCACTC <u>GGAGG</u> ACTTAACAC	-6,856025	ATG	UGL61140.1	internal virion protein
#39	+	27423	29675	750	AAACTTCTGT <u>AGGAGGT</u> AAC	-11,181535	ATG	UGL61141.1	internal virion protein
#40	+	29684	33511	1275	CCTAATAAGT <u>TGAGGG</u> CCACA	-5,05103	ATG	UGL61142.1	internal virion protein
#41	+	33572	35464	630	AACCAATTCT <u>GGAGA</u> ACCAA	-3,60793	ATG	UGL61143.1	tail fiber protein
#42	+	35474	35818	114	GAGTTTCGT <u>TAAGGAG</u> ATACA	-7,94194	ATG	UGL61144.1	putative tail fiber assembly protein
#43	+	35815	35997	60	CAAATCA <u>AGGTGAT</u> GCAAGA	-9,054465	ATG	UGL61145.1	holin
#44	+	35984	36286	100	CTACGAGA <u>AGGT</u> CAAGAGCA	-3,49838	ATG	UGL61146.1	terminase small subunit
#45	+	36286	38064	592	GCCCAAAT <u>GGGT</u> ATTCACTG	-3,546615	ATG	UGL61147.1	terminase large subunit
#46	+	38068	38328	86	CTCGTAA <u>AAGGA</u> AATGATCC	-4,53508	ATG	UGL61148.1	hypothetical protein
#47	+	38340	38639	99	GACCTGT <u>TAAGGAG</u> CAACACC	-7,94194	ATG	UGL61149.1	hypothetical protein
#48	+	38643	39143	166	ACGCT <u>GTG</u> CTGTCCTAAGTC	-0,272315	ATG	UGL61150.1	endolysin
#49	+	39140	39445	101	CTCTGG <u>AAGGAGG</u> TTATCA	-9,85811	ATG	UGL61151.1	putative i-spanin
#50	+	39354	39584	76	CACCCGC <u>AAGGAGGTGA</u> GCG	-16,565775	ATG	UGL61152.1	putative o-spanin
#51	+	39675	40163	162	AGGTCACCT <u>AGGA</u> CAAAAAA	-3,60793	CTG	UGL61153.1	hypothetical protein

Supplementary Figure S1. Viridic heatmap showing pairwise intergenomic distances (%) of *Pseudomonas* phages Eir4 and Eisa9 with other related phages. Labels of *Pseudomonas* infecting *Autographiviridae* phages with a standing in the official phage taxonomy (subset from the ICTV virus metadata resource; n=49) are in black font color. Labels of ten phages most similar to Eir4 that were not found as a part of *Pseudomonas*-infecting *Autographiviridae* phage VMR subset are in red font color (n=10), label of Eir4 is in red and is also made bold. Labels of ten phages most similar to Eisa9 that were not found as a part of *Pseudomonas*-infecting *Autographiviridae* phage VMR subset are in blue font color (n=10), label of Eisa9 is also made bold. Red rectangle indicates a cluster of phage genomes that are either officially recognized as representatives of genus *Ghuvavirus* or can be proposed for inclusion in it. Maroon rectangle indicates *Pseudomonas* phage isolate genomes that belong to the yet unproposed same phage species as Eir4. Purple rectangle indicates *Pseudomonas* phages PollyC and Eisa9, which can probably be proposed as an exemplar isolate for the creation of a novel species in the *Polyceevirus* genus.



Supplementary Figure S2. Terminase large subunit protein amino acid sequence maximum-likelihood tree for prediction of packaging strategy employed by phages Eir4 and Eisa9. The tree is drawn to scale, branch lengths correspond to the number of amino acid differences per site. Tips are labeled as “protein accession | originating phage | packaging strategy/genome termini type employed by the phage”, labels of the tips representing the amino acid sequences of Eir4 and Eisa9 proteins are coloured in red and blue, respectively. The tree is midpoint rooted and the distal nodes of branches with $\geq 95\%$ UFBoot support (out of 1000 replicates) are indicated by green squares. Maroon rectangle indicates the clade containing MRCA of T7 type SDTR packaging strategy employing phages and its children. “Unknown*” indicates that the TerL sequence comes from a phage for which the packaging strategy is unknown. LDTR stands for long direct terminal repeats, SDTR - for short direct terminal repeats.



Supplementary Figure S3. Pairwise genome nucleotide sequence comparison of *Autographiviridae* phage T7 as an exemplar *Autographiviridae* phage, phage 17A (representative of *Ghunavirus* genus) as the closest Eir4 relative with the standing in the current phage taxonomy, *Pseudomonas* phage Eir4 itself and eight phages that belong to the same putative novel species in genus *Ghunavirus* as Eir4 using BLASTN. Genome representations are drawn to scale with the scale bar indicating 2000 base pairs. Arrows represent open reading frames and point in the direction of transcription. Color-coding is based on the functional groups of the respective ORF putative products according to the legend and ORFs of Eir4 with functional prediction are annotated by the order of their appearance in the genome (Supplementary tables 1 and 2). Gray boxes represent regions of similarity between the genomes and are coloured in gradient with darker shades of gray representing higher region identity. The figure was generated using EasyFig (Sullivan et al., 2011).

