	a	b	c
Prediction exactly matches GenBank annotation	201	215	220
Prediction differs in start location from GenBank annotation	52	34	31
Prediction overlaps with GenBank annotated interrupted gene	9	9	9
Prediction does not match any gene annotated in GenBank	9	6	6
GenBank annotated gene was not predicted	12	16	14

Web Supplement Table 1. Comparison of results of analysis of the phage T4 genome using the *ab initio* gene finding method with different settings: (a) use of heuristic models only; (b) use of GeneMarkS with RBS derivation off; (c) use of GeneMarkS with RBS derivation on.

	HHV1	HHV2	HHV3	HHV4	HHV5	HHV6A	HHV6B	HHV7	HHV8
Start prediction is confirmed by BLASTP									
start prediction agrees with annotation	47	53	37	37	48	48	39	45	50
possibly wrong annotation	3	2	2	0	9	5	5	1	5
BLASTP indicates longer protein									
start prediction agrees with annotation	0	0	2	5	1	3	1	9	2
possibly wrong prediction	6	2	1	2	10	2	2	3	4
BLASTP indicates shorter protein									
start prediction agrees with annotation	2	0	3	1	2	4	3	1	1
possibly wrong prediction	0	0	0	0	0	1	0	0	1

Web Supplement Table 2. Comparison of predicted and annotated starts for genes from nine human herpesviruses with BLASTP hits to non-redundant database. The Kozak translational start model was used for predictions.

Accession #_strand_lend_rend	E-value	GenBank definition
NC_000852_c_1174_1335		Paramecium bursaria Chlorella virus 1, complete genome.
NC_000852_d_329409_329570	3.00E-18	Paramecium bursaria Chlorella virus 1, complete genome.
NC_000852_c_329231_329377		Paramecium bursaria Chlorella virus 1, complete genome.
NC_000852_d_1367_1513	3.00E-26	Paramecium bursaria Chlorella virus 1, complete genome.
NC_000900_c_155390_155554		Variola minor virus, complete genome.
NC_001611_c_154328_154492	6.00E-28	Variola virus, complete genome.
NC_000900_c_161641_161796		Variola minor virus, complete genome.
NC_001611_c_160577_160732	7.00E-24	Variola virus, complete genome.
NC_000900_c_181944_182048		Variola minor virus, complete genome.
NC_001611_c_180216_180320	9.00E-16	Variola virus, complete genome.
NC_000900_c_18713_18856		Variola minor virus, complete genome.
NC_001611_c_17887_18030	1.00E-22	Variola virus, complete genome.
NC_000900_c_24275_24388		Variola minor virus, complete genome.
NC_001611_c_23277_23390	5.00E-18	Variola virus, complete genome.
NC_000900_c_34087_34236		Variola minor virus, complete genome.
NC_001611_c_33100_33249	4.00E-22	Variola virus, complete genome.
NC_003391_c_45121_45270	3.00E-21	Camelpox virus, complete genome.
NC_003663_c_58921_59070	2.00E-20	Cowpox virus, complete genome.
NC_003310_c_42514_42660	1.00E-19	Monkeypox virus, complete genome.
NC_002170_c_41808_41954	7.00E-19	Vaccinia virus (strain Tian Tan), complete genome.
NC_000900_c_6726_6902		Variola minor virus, complete genome.
NC_001611_c_6694_6870	9.00E-07	Variola virus, complete genome.
NC_000900_c_85_258		Variola minor virus, complete genome.
NC_001611_c_51_224	2.00E-31	Variola virus, complete genome.
NC_000900_d_136580_136732		Variola minor virus, complete genome.
NC_001611_d_135556_135708	3.00E-26	Variola virus, complete genome.
NC_000900_d_159297_159470		Variola minor virus, complete genome.
NC_001611_d_158234_158407	3.00E-29	Variola virus, complete genome.
NC_000900_d_161330_161470		Variola minor virus, complete genome.
NC_001611_d_160266_160406	8.00E-23	Variola virus, complete genome.
NC_000900_d_163206_163355		Variola minor virus, complete genome.
NC_001611_d_161545_161694	2.00E-24	Variola virus, complete genome.
NC_000900_d_175689_175838		Variola minor virus, complete genome.
NC_001611_d_173961_174110	2.00E-26	Variola virus, complete genome.
NC_000924_d_27423_27569		Bacteriophage 933W provirus, complete genome.

NC_003525_d_59172_59321	2.00E-20	Stx2 converting bacteriophage I virion, complete genome.
NC_000924_d_57568_57663		Bacteriophage 933W provirus, complete genome.
NC_003525_d_27653_27748	1.00E-13	Stx2 converting bacteriophage I virion, complete genome.
NC_001132_c_161090_161194		Myxoma virus, complete genome.
NC_001132_d_580_684	2.00E-16	Myxoma virus, complete genome.
NC_001132_c_23703_23834		Myxoma virus, complete genome.
NC_001266_c_22906_23037	3.00E-15	Rabbit fibroma virus, complete genome.
NC_001132_d_102186_102290		Myxoma virus, complete genome.
NC_001266_d_101359_101463	9.00E-13	Rabbit fibroma virus, complete genome.
NC_001266_c_149358_149630		Rabbit fibroma virus, complete genome.
NC_001266_d_10228_10500	6.00E-50	Rabbit fibroma virus, complete genome.
NC_001266_c_149611_149784		Rabbit fibroma virus, complete genome.
NC_001266_d_10074_10247	1.00E-23	Rabbit fibroma virus, complete genome.
NC_001266_c_157852_158031		Rabbit fibroma virus, complete genome.
NC_001266_d_1827_2006	2.00E-31	Rabbit fibroma virus, complete genome.
NC_001266_c_158342_158401		Rabbit fibroma virus, complete genome.
NC_001266_d_1457_1516	7.00E-06	Rabbit fibroma virus, complete genome.
NC_001266_c_2085_2276		Rabbit fibroma virus, complete genome.
NC_001266_d_157582_157773	1.00E-34	Rabbit fibroma virus, complete genome.
NC_001330_d_206_433		Coliphage alpha3, complete genome.
NC_001730_d_4_432	3.00E-23	Coliphage phiK, complete genome.
NC_001355_c_4005_4124		Human papillomavirus type 6b, complete genome.
NC_001668_c_4006_4125	2.00E-16	Human papillomavirus type 6a, complete genome.
NC_001406_c_22443_24032		Human adenovirus type 5, complete genome.
NC_001997_c_22668_24041	3.00E-87	Porcine adenovirus A, complete genome.
NC_001406_c_30864_31031		Human adenovirus type 5, complete genome.
NC_003266_c_31966_32121	1.00E-15	Human adenovirus E, complete genome.
NC_001406_c_35141_35323		Human adenovirus type 5, complete genome.
NC_003266_c_35734_36108	4.00E-10	Human adenovirus E, complete genome.
NC_001406_c_5197_8367		Human adenovirus type 5, complete genome.
NC_003266_c_6073_8454	0	Human adenovirus E, complete genome.
NC_001997_c_5184_7826	0	Porcine adenovirus A, complete genome.
NC_001406_c_8583_10544		Human adenovirus type 5, complete genome.
NC_003266_c_8457_9440	9.00E-09	Human adenovirus E, complete genome.
NC_001997_c_8555_10309	4.00E-97	Porcine adenovirus A, complete genome.

NC_001406_d_2019_3509		Human adenovirus type 5, complete genome.
NC_001997_d_1829_3253	5.00E-51	Porcine adenovirus A, complete genome.
NC_001406_d_21733_22347		Human adenovirus type 5, complete genome.
NC_003266_d_21527_21763	2.00E-28	Human adenovirus E, complete genome.
NC_003266_d_21144_21329	2.00E-28	Human adenovirus E, complete genome.
NC_001406_d_3609_4031		Human adenovirus type 5, complete genome.
NC_001997_d_3256_3993	1.00E-06	Porcine adenovirus A, complete genome.
NC_001406_d_560_1120		Human adenovirus type 5, complete genome.
NC_001997_d_533_1222	3.00E-11	Porcine adenovirus A, complete genome.
NC_001426_d_2_109		Enterobacteria phage GA, complete genome.
NC_002250_d_3_110	7.00E-07	Enterobacteria phage KU1, complete genome.
NC_001493_c_119119_119211		Ictalurid herpesvirus 1, complete genome.
NC_001493_c_3449_3541	3.00E-13	Ictalurid herpesvirus 1, complete genome.
NC_001493_c_121535_121756		Ictalurid herpesvirus 1, complete genome.
NC_001493_c_5865_6086	4.00E-40	Ictalurid herpesvirus 1, complete genome.
NC_001493_c_123527_123640		Ictalurid herpesvirus 1, complete genome.
NC_001493_c_7857_7970	1.00E-17	Ictalurid herpesvirus 1, complete genome.
NC_001493_d_10517_10759		Ictalurid herpesvirus 1, complete genome.
NC_001493_d_126187_126429	1.00E-46	Ictalurid herpesvirus 1, complete genome.
NC_001501_c_23_307		Murine leukemia virus, complete genome.
NC_001819_c_120_797	4.00E-44	Rauscher murine leukemia virus, complete genome.
NC_002603_c_302_805	2.00E-22	DG-75 Murine leukemia virus, complete genome.
NC_001511_d_2121_5051		Ovine lentivirus, complete genome.
NC_001702_d_2729_5845	4.00E-17	Murine type C retrovirus, complete genome.
NC_001548_d_2467_3219	7.00E-09	Simian SRV-2 type D retrovirus, complete genome.
NC_003059_d_5534_6628	3.00E-06	Porcine endogenous retrovirus, complete genome.
NC_003059_d_3307_3840	8.00E-06	Porcine endogenous retrovirus, complete genome.
NC_001514_d_24_152		Woolly monkey sarcoma virus, complete genome.
NC_001514_d_8329_8481	3.00E-23	Woolly monkey sarcoma virus, complete genome.
NC_001544_d_81_5669		Ross River virus, complete genome.
NC_002633_d_63_3563	1.00E-11	Cucumber fruit mottle mosaic virus, complete genome.
NC_001550_c_7723_7998		Mason-Pfizer monkey virus, complete genome.
NC_001550_c_8080_8340	5.00E-38	Mason-Pfizer monkey virus, complete genome.
NC_001577_d_3201_3956		Human papillomavirus type 12, complete genome.
NC_001578_d_2851_3696	2.00E-28	Human papillomavirus type 14, complete genome.
NC_001580_d_3196_3786	6.00E-14	Human papillomavirus type 17, complete genome.
NC_001579_d_3072_3770	1.00E-12	Human papillomavirus type 15, complete genome.
NC_001596_d_3199_3816	9.00E-10	Human papillomavirus type 9, complete genome.

NC_001591_d_3109_3903	2.00E-08	Human papillomavirus type 49, complete genome.
NC_001578_d_121_744		Human papillomavirus type 14, complete genome.
NC_001581_d_200_700	4.00E-46	Human papillomavirus type 19, complete genome.
NC_001582_d_3087_3185		Human papillomavirus type 25, complete genome.
NC_001679_d_3112_3312	7.00E-08	Human papillomavirus type 20, complete genome.
NC_001588_c_7686_7793		Human papillomavirus type 3, complete genome.
NC_001684_c_7835_7918	1.00E-10	Human papillomavirus type 28, complete genome.
NC_001590_c_4445_4864		Human papillomavirus type 45, complete genome.
NC_001692_c_4450_4965	2.00E-14	Human papillomavirus type 55, complete genome.
NC_001596_d_2809_3033		Human papillomavirus type 9, complete genome.
NC_001688_d_2800_2913	4.00E-06	Human papillomavirus type 38, complete genome.
NC_001600_d_38_781		Cardamine chlorotic fleck virus, complete genome.
NC_001777_d_60_668	3.00E-08	Tobacco necrosis virus A, complete genome.
NC_001611_d_149894_150058		Variola virus, complete genome.
NC_002170_d_158432_159004	9.00E-26	Vaccinia virus (strain Tian Tan), complete genome.
NC_003310_d_157837_157971	4.00E-18	Monkeypox virus, complete genome.
NC_001623_c_61009_61083		Autographa californica nucleopolyhedrovirus, complete genome.
NC_001962_c_55370_55444	3.00E-08	Bombyx mori nuclear polyhedrosis virus, complete genome.
NC_001639_d_11006_11218		Lactate dehydrogenase-elevating virus, complete genome.
NC_002534_d_11074_11286	5.00E-32	Lactate dehydrogenase-elevating virus, complete genome.
NC_001961_d_12078_12299	2.00E-15	Porcine reproductive and respiratory syndrome virus, complete
NC_002533_d_11801_12013	3.00E-14	Lelystad virus, complete genome.
NC_001650_d_15068_15259		Equine herpesvirus 2, complete genome.
NC_001650_d_181942_182133	6.00E-37	Equine herpesvirus 2, complete genome.
NC_001707_d_295_1497		Hepatitis B virus, complete genome.
NC_001896_d_423_1637	0	Woolly monkey hepatitis B Virus, complete genome.
NC_001734_c_25100_25384		Canine adenovirus, complete genome.
NC_001784_c_25289_25567	4.00E-26	Canine adenovirus type 2, complete genome.
NC_001781_c_4752_4877		Human respiratory syncytial virus, complete genome.
NC_001803_c_4721_4999	1.00E-11	Respiratory syncytial virus, complete genome.
NC_001792_d_1170_1457		Porcine circovirus, complete genome.
NC_002173_d_19_366	8.00E-08	Porcine circovirus type 2, complete genome.
NC_001825_c_107_232		Streptococcus phage Cp-1, complete genome.
NC_001825_d_19112_19237	2.00E-19	Streptococcus phage Cp-1, complete genome.
NC_001844_c_140003_140200		Equine herpesvirus 4, complete genome.
NC_001844_d_117850_118047	2.00E-36	Equine herpesvirus 4, complete genome.

NC_001844_c_140498_140629		Equine herpesvirus 4, complete genome.
NC_001844_d_117421_117552	9.00E-22	Equine herpesvirus 4, complete genome.
NC_001899_c_157_432		Diatraea saccharalis densovirus, complete genome.
NC_001899_d_5577_5852	7.00E-44	Diatraea saccharalis densovirus, complete genome.
NC_001933_c_2280_2540		Indian cassava mosaic virus DNA-B, complete sequence.
NC_003862_c_2161_2544	6.00E-46	Sri Lankan cassava mosaic virus DNA B, complete sequence.
NC_001940_d_685_2427		Feline leukemia virus, complete genome.
NC_002603_d_627_2237	2.00E-39	DG-75 Murine leukemia virus, complete genome.
NC_003059_d_2375_2689	2.00E-09	Porcine endogenous retrovirus, complete genome.
NC_001955_c_9352_9906		Bacteriophage phi CTX, complete genome.
NC_003278_c_9352_9906	2.00E-74	Bacteriophage phi CTX, complete genome.
NC_001955_d_26237_26539		Bacteriophage phi CTX, complete genome.
NC_003278_d_26237_26539	2.00E-54	Bacteriophage phi CTX, complete genome.
NC_001955_d_27134_27676		Bacteriophage phi CTX, complete genome.
NC_003278_d_27134_27676	2.00E-90	Bacteriophage phi CTX, complete genome.
NC_001973_c_54400_54636		Lymantria dispar nucleopolyhedrovirus, complete genome.
NC_002169_d_102482_102685	1.00E-11	Spodoptera exigua nucleopolyhedrovirus, complete genome.
NC_001997_d_11683_13587		Porcine adenovirus A, complete genome.
NC_003266_d_12024_13130	2.00E-97	Human adenovirus E, complete genome.
NC_003266_d_13094_13804	4.00E-18	Human adenovirus E, complete genome.
NC_001997_d_17783_18001		Porcine adenovirus A, complete genome.
NC_003266_d_17261_17410	7.00E-08	Human adenovirus E, complete genome.
NC_002170_c_15769_15858		Vaccinia virus (strain Tian Tan), complete genome.
NC_003663_c_32729_32818	2.00E-12	Cowpox virus, complete genome.
NC_002170_c_186115_186174		Vaccinia virus (strain Tian Tan), complete genome.
NC_002170_d_3101_3160	1.00E-06	Vaccinia virus (strain Tian Tan), complete genome.
NC_002170_c_21836_22126		Vaccinia virus (strain Tian Tan), complete genome.
NC_003391_c_25103_25243	1.00E-18	Camelpox virus, complete genome.
NC_002170_c_28451_28585		Vaccinia virus (strain Tian Tan), complete genome.
NC_003391_c_31678_31812	1.00E-22	Camelpox virus, complete genome.
NC_002170_c_5150_5353		Vaccinia virus (strain Tian Tan), complete genome.
NC_002170_d_183934_184125	5.00E-35	Vaccinia virus (strain Tian Tan), complete genome.
NC_003391_c_10696_10869	4.00E-10	Camelpox virus, complete genome.
NC_003310_d_187266_187403	2.00E-06	Monkeypox virus, complete genome.
NC_002170_d_146345_146539		Vaccinia virus (strain Tian Tan), complete genome.
NC_003310_d_146245_146424	5.00E-16	Monkeypox virus, complete genome.

NC_002321_c_30318_30428		Bacteriophage phi PVL provirus, complete genome.
NC_003288_c_6701_6811	4.00E-16	Bacteriophage phi ETA, complete genome.
NC_002362_d_4858_5001		Bacteriophage VfO3K6, complete genome.
NC_002363_d_4858_5001	5.00E-23	Bacteriophage VfO4K68, complete genome.
NC_002499_d_2946_3158		Human astrovirus type 8, complete genome.
NC_002597_d_2945_3157	2.00E-26	Human astrovirus strain Oxford.
NC_002666_d_2_64		Bacteriophage bIL285, complete genome.
NC_002703_d_2_64	8.00E-07	Bacteriophage Tuc2009, complete genome.
NC_002530_d_22925_23206		Gallid herpesvirus 1 (serotype 2), complete genome.
NC_002577_d_34749_35030	5.00E-49	Gallid herpesvirus 3, complete genome.
NC_002577_c_126686_127000		Gallid herpesvirus 3, complete genome.
NC_002577_d_6583_6897	1.00E-59	Gallid herpesvirus 3, complete genome.
NC_002577_c_129283_129486		Gallid herpesvirus 3, complete genome.
NC_002577_d_4097_4300	5.00E-30	Gallid herpesvirus 3, complete genome.
NC_002577_c_130747_131019		Gallid herpesvirus 3, complete genome.
NC_002577_d_2564_2836	2.00E-52	Gallid herpesvirus 3, complete genome.
NC_002577_c_141755_141961		Gallid herpesvirus 3, complete genome.
NC_002577_d_155510_155716	3.00E-37	Gallid herpesvirus 3, complete genome.
NC_002577_c_156035_156250		Gallid herpesvirus 3, complete genome.
NC_002577_d_141221_141436	3.00E-23	Gallid herpesvirus 3, complete genome.
NC_002577_c_3857_4012		Gallid herpesvirus 3, complete genome.
NC_002577_d_129571_129726	3.00E-27	Gallid herpesvirus 3, complete genome.
NC_002577_c_4709_5026		Gallid herpesvirus 3, complete genome.
NC_002577_d_128557_128874	2.00E-46	Gallid herpesvirus 3, complete genome.
NC_003027_c_112394_112519		Lumpy skin disease virus, complete genome.
NC_003389_c_110617_110742	5.00E-07	Swinepox virus, complete genome.
NC_003094_c_28392_28514		Helicoverpa armigera nuclear polyhedrosis virus, complete genome.
NC_003349_c_28207_28329	4.00E-20	Helicoverpa zea single nucleocapsid nucleopolyhedrovirus, complete
NC_003094_d_110418_110561		Helicoverpa armigera nuclear polyhedrosis virus, complete genome.
NC_003349_d_110540_110683	2.00E-24	Helicoverpa zea single nucleocapsid nucleopolyhedrovirus, complete
NC_003310_c_135169_135333		Monkeypox virus, complete genome.
NC_003391_c_138073_138228	3.00E-07	Camelpox virus, complete genome.
NC_003310_c_141484_141612		Monkeypox virus, complete genome.
NC_003663_c_158938_159066	5.00E-18	Cowpox virus, complete genome.

NC_003391_c_144185_144313	2.00E-17	Camelpox virus, complete genome.
NC_003310_c_196261_196386		Monkeypox virus, complete genome.
NC_003310_d_473_598	5.00E-20	Monkeypox virus, complete genome.
NC_003663_c_222481_222594	4.00E-13	Cowpox virus, complete genome.
NC_003663_d_1908_2021	4.00E-13	Cowpox virus, complete genome.
NC_003345_c_35_256		Halovirus HF2, complete genome.
NC_003345_c_77399_77620	2.00E-22	Halovirus HF2, complete genome.
NC_003389_c_145022_145081		Swinepox virus, complete genome.
NC_003389_d_1374_1433	4.00E-06	Swinepox virus, complete genome.
NC_003391_d_151253_151330		Camelpox virus, complete genome.
NC_003663_d_166025_166507	6.00E-07	Cowpox virus, complete genome.
NC_003407_c_315_1028		Rana tigrina ranavirus, complete genome.
NC_003494_c_83701_84663	1.00E-36	Infectious spleen and kidney necrosis virus, complete genome.
NC_003407_c_52315_52899		Rana tigrina ranavirus, complete genome.
NC_003494_c_81707_82177	9.00E-12	Infectious spleen and kidney necrosis virus, complete genome.
NC_003407_d_22319_22774		Rana tigrina ranavirus, complete genome.
NC_003494_c_110285_110656	2.00E-07	Infectious spleen and kidney necrosis virus, complete genome.
NC_003407_d_27093_28241		Rana tigrina ranavirus, complete genome.
NC_003407_d_28636_29733	9.00E-43	Rana tigrina ranavirus, complete genome.
NC_003839_d_7042_7269		Tomato ringspot virus RNA 2, complete sequence.
NC_003840_d_7983_8210	2.00E-32	Tomato ringspot virus RNA 1, complete sequence.
NC_003867_c_80_136		Havana tomato virus DNA A, complete sequence.
NC_003868_c_81_251	9.00E-06	Havana tomato virus DNA B, complete sequence.

Web Supplement Table 3. List of new predictions with BLASTP hits to other new predictions as clustered into unique groups. The format of the left column refers to the accession number of the genome, the strand the gene was found in (d for direct, c for complementary), and the coordinates of the gene (left and right ends). The query protein for each cluster is shown in bold. The middle column shows the E-value for each hit. The right column gives the GenBank definition for the viral genome.