

SUPPLEMENTAL MATERIAL

A cloud-compatible bioinformatics pipeline for ultra-rapid pathogen identification from next-generation sequencing of clinical samples

Samia N. Naccache^{1,2}, Scot Federman^{1,2}, Narayanan Veerarghavan^{1,2}, Matei Zaharia³, Deanna Lee^{1,2}, Erik Samayoa^{1,2}, Jerome Bouquet^{1,2}, Alexander L. Greninger⁴, Ka-Cheung Luk⁵, Barryett Enge⁶, Debra A. Wadford⁶, Sharon L. Messenger⁶, Gillian L. Genrich¹, Kristen Pellegrino⁷, Gilda Grard⁸, Eric Leroy⁸, Bradley S. Schneider⁹, Joseph N. Fair⁹, Miguel A. Martinez¹⁰, Pavel Isa¹⁰, John A. Crump^{11,12,13}, Joseph L. DeRisi⁴, Taylor Sittler¹, John Hackett, Jr.⁵, Steve Miller^{1,2} and Charles Y. Chiu^{1,2,14*}

Affiliations:

¹Department of Laboratory Medicine, UCSF, San Francisco, CA, USA

²UCSF-Abbott Viral Diagnostics and Discovery Center, San Francisco, CA, USA

³Department of Computer Science, University of California, Berkeley, CA, USA

⁴Department of Biochemistry, UCSF, San Francisco, CA, USA

⁵Abbott Diagnostics, Abbott Park, IL, USA

⁶Viral and Rickettsial Disease Laboratory, California Department of Public Health, Richmond, CA, USA

⁷Department of Family and Community Medicine, UCSF, San Francisco, CA, USA

⁸Viral Emergent Diseases Unit, Centre International de Recherches Médicales de Franceville, Franceville, Gabon

⁹Metabiota, Inc., San Francisco, CA, USA

¹⁰Departamento de Genética del Desarrollo y Fisiología Molecular, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, México

¹¹Division of Infectious Diseases and International Health and the Duke Global Health Institute, Duke University Medical Center, Durham, NC

¹²Kilimanjaro Christian Medical Centre, Moshi, Tanzania

¹³Centre for International Health, University of Otago, Dunedin, New Zealand

¹⁴Department of Medicine, Division of Infectious Diseases, UCSF, San Francisco, CA, USA

*Corresponding Author

Charles Chiu

Department of Laboratory Medicine

UCSF, San Francisco, CA

charles.chiu@ucsf.edu

TABLE OF CONTENTS

Supplemental Results.....	3
Supplemental Methods.....	4
Supplemental Figures.....	15
Supplemental Figure S1. Effect of read length on accuracy of SNAP and RAPSearch classification.....	15
Supplemental Figure S2. Coverage maps automatically generated by SURPI in comprehensive mode.....	16
Supplemental Figure S3. Relative percentages of processing time and mapped / filtered reads for clinical metagenomic datasets analyzed using SURPI.....	21
Supplemental Figure S4. ROC curves generated from varying BWA and BT2 parameter combinations.....	22
Supplemental Figure S5. SURPI scripts and software / database dependencies.....	23
Supplemental Tables.....	24
Supplemental Table S1. Processing time and cost of SURPI versus PathSeq (v1.2) on a cloud server.....	24
Supplemental Table S2. Composition and number of reads corresponding to <i>in silico</i> -generated NGS datasets.....	25
Supplemental Table S3. Clinical NGS datasets used for ROC curve analysis.....	26
Supplemental Table S4. Clinical NGS datasets used for SURPI performance validation.....	27
Supplemental Table S5. Sensitivity and specificity of viral detection in clinical NGS datasets.....	28
Supplemental Tables S6-S13. Viral species summary tables of SURPI (SNAP) nucleotide alignment to the NCBI nt DB.....	29
Supplemental Tables S14-S21. Viral species summary tables of SURPI (RAPSearch) translated nucleotide alignment to the viral protein DB.....	48
Supplemental Table S22. Performance testing of <i>de novo</i> assembly methods.....	71
Supplemental References.....	72

SUPPLEMENTAL RESULTS

Effect of read length on SNAP and RAPSearch performance

The ability of SNAP and RAPSearch to correctly identify viral reads across varying read lengths was evaluated (Supplemental Fig. S1A-B). Since SNAP is a global aligner, the optimal edit distance determined by Youden's index and the F_1 score (harmonic mean) (Akobeng 2007) increases linearly as the read length increases (Supplemental Fig. S1C). For RAPSearch, the optimal E-value cutoff decreases exponentially as the read length increases (Supplemental Fig. S1D).

Comparison of processing time and costs for running SURPI versus PathSeq v2.1 on a cloud server.

Both SURPI and PathSeq v2.1 can be run in a cloud computing environment using an Amazon Elastic Compute Cloud (EC2) server. We benchmarked SURPI using a hi1.4xlarge instance with 16vCPUs, 60.5 GB RAM, and 2 x 1024 GB SSDs. PathSeq was run using its default "cloud" setup and with default parameters on EC2, using a Hadoop cluster of 20 m1.large instances as described (Kostic et al. 2011). Overall processing times and costs are reported in Supplemental Table S1. Costs were calculated by hour (minutes rounded up) according to established Amazon EC2 prices as of July of 2013.

Accuracy of SURPI for viral read identification in clinical NGS datasets

SURPI's ability to correctly identify viral reads originating from viral genomes known to be represented in clinical NGS datasets was evaluated. Clinical sample sets benchmarked and validated by SURPI are shown in Table S4. True viral reads / microbial reads were defined as meeting an E-value cutoff of 1×10^{-20} after BLASTn alignment to the most closely related viral genome in the reference database. Using this definition, the sensitivity and specificity of SURPI for viral identification were calculated (Supplemental Table S5). For detection of parechovirus in pediatric stool, there is decreased sensitivity by SURPI in fast mode due to sequence divergence (Figure 5B; Supplemental Table S4), as fast mode, unlike comprehensive mode, does not use RAPSearch for translated nucleotide alignment to the viral database. Similarly, for the highly divergent titi monkey adenovirus (TMAdV) and Bas-Congo rhabdovirus (BASV), only 8% and 0% viral reads are detected by nucleotide alignment, respectively, using SNAP, while only 14-16% viral reads are detected using RAPSearch (Figure 5G; Supplemental Table S4).

SUPPLEMENTAL METHODS

Clinical samples

NGS datasets from clinical samples harboring defined pathogens were generated in-house with the exception of datasets corresponding to a prostate cancer cell line (Prensner et al. 2011) and biopsy tissue from a colon cancer study (Castellarin et al. 2012), which were downloaded from the publicly available NCBI Sequence Read Archive (SRA). Details of NGS library preparation have been previously reported for the following datasets: chronic hepatitis (serum) (Naccache et al. 2013), pediatric diarrhea (stool) (Yu et al. 2012), hemorrhagic fever (serum) (Grard et al. 2012), prostate cancer (biopsy tissue) (Lee et al. 2012), simian pneumonia (lung swab / tissue) (Chen et al. 2011), and acute respiratory illness from influenza A(H1N1)pdm09 (nasal swab) (Greninger et al. 2010) (Supplementary Methods). The remaining in-house NGS libraries were derived from the following diseases and clinical samples: transfusion-transmitted viral hepatitis from the transfusion-transmitted viruses study (TTVS) (serum), (Mosley et al. 1990), hantavirus pulmonary syndrome (serum) (Hartline et al. 2013), encephalitis (cerebrospinal fluid), spiked HIV (plasma), acute hemorrhagic fever outbreak in the Democratic Republic of the Congo, Africa (serum), acute febrile illness in Tanzania, Africa (serum), and fever in a returning traveler (serum).

Ethics statement

All clinical samples were analyzed under protocols approved by the UCSF Institutional Review Board (IRB) and the IRBs of the collaborative donor institutions. Written informed consent to provide clinical samples for microbial analysis was previously obtained for patients in the chronic hepatitis cohort (University of Chicago) (Naccache et al. 2013) and the BIOLINCC Transfusion-Transmitted Viruses Study (Mosley et al. 1990), and for the returning traveler with undiagnosed fever. Other samples did not require consent as they were pre-existing and de-identified or were obtained from public health investigations, thus deemed not to constitute human subjects research

NGS library construction

The majority of NGS libraries were constructed using a modified TruSeq (Illumina, San Diego, CA) protocol as previously described (Naccache et al. 2013). Briefly, reverse transcription was performed using random nonamers attached to a 18 bp linker adapter (Wang et al. 2003), followed by second-strand synthesis using Sequenase (Affymetrix, Santa Clara,

CA) and 10-25 cycles of PCR amplification using high-fidelity KlenTaq LA DNA polymerase (Sigma-Aldrich, Carlsbad, CA). Adapter sequences were then cleaved using the Type IIs restriction endonuclease *GsuI* or *BpmI*, followed by end-repair, A-tailing, truncated Illumina adapter ligation, and size selection using a 2% agarose gel or AMPure XP beads (New England Biolabs, Ipswich, MA), targeting an insert size of 270–350 bp. Full-length Illumina adapters containing 6-bp barcodes were then added via an additional 17 cycles of PCR amplification. Libraries were validated using the Kapa qPCR kit (Kapa Biosystems) and Agilent Bioanalyzer (Agilent, Santa Clara, CA), and then sent to an outside company (Elim Biopharmaceuticals, Hayward, CA) for NGS on an Illumina GAIIx or HiSeq 2000 or run in-house on an Illumina MiSeq instrument.

NGS datasets (Figs. 5-7; SupplementalTable S4)

A. Chronic hepatitis (serum). As previously described (Naccache et al. 2013), a total of 169 serum samples from a clinical study of patients with seronegative hepatitis were collected at the University of Chicago as part of a chronic hepatitis study. Patient samples (n=105) were combined into 9 pools, passed through a 0.22 μ m filter (Millipore, Hayward, CA), and pretreated with nucleases as described (Swei et al. 2013), after which nucleic acids were extracted using the EZ1 Viral Mini Kit v2.0 (Qiagen, Carlsbad, CA), and Illumina TruSeq-adapted libraries were prepared using an Eppendorf epMotion 5075 BioRobot. A total of 508,729,590 million reads obtained from sequencing 2 lanes of an Illumina HiSeq 2000 100 bp paired-end (PE) run were analyzed.

B. Pediatric diarrhea (stool). As previously reported (Yu et al. 2012), anonymized stool samples were collected from 96 children with acute diarrheal disease from 3 different states in Mexico between 2008–2009. Viral particles were purified from stool samples by fecal preparation in chloroform, followed by filtering and treatment with a nuclease cocktail, prior to nucleic acid extraction and NGS library generation. A total of 158,026,920 reads corresponding to a pool of 16 samples were obtained from a HiSeq 2000 100 bp PE run. Analysis of the NGS data led to the discovery of a novel polyomavirus (Yu et al. 2012), and detection of numerous additional viruses, including a human parechovirus 1 (HPeV-1 isolate MX1, GenBank accession number KJ152442), sapovirus (SaV), rotavirus, astrovirus, respiratory syncytial virus (RSV), rhinovirus, and norovirus (Figure 5B). The detection of each of these viruses by NGS was subsequently confirmed using specific PCR (Supplemental Table S4).

C. Colon cancer (biopsy tissue). NGS data corresponding to 11 colorectal tumor samples and 11 normal tissue samples was retrieved from the NCBI Sequence Read Archive

(SRA accession number SRP007584). The total dataset consisted of 143,419,974 84 bp reads and was previously found to contain a greater proportion of *Fusobacterium nucleatum* reads in the colorectal tumor samples (Castellarin et al. 2012). Normalization of bacterial read abundance was performed using Cluster 3.0 (de Hoon et al. 2004).

D and H. Hemorrhagic fever (serum). As previously reported (Grard et al. 2012), NGS enabled detection of the novel rhabdovirus Bas-Congo virus (BASV) in a serum sample from a patient from central Africa with acute hemorrhagic fever. A total of 100 million reads from the published NGS dataset were selected for analysis using SURPI (Fig. 5G, upper panel; Fig. 7; Supplemental Table 4, “D”). RNA extracted from a different aliquot of the same patient serum sample was converted into cDNA and then used to generate an independent NGS library (Fig. 5G, lower panel; Fig. 7; Supplemental Table 4, “H”). For the second library, 17,881,580 reads obtained from a 250 bp PE Illumina MiSeq run were analyzed.

E. TTVS hepatitis (serum). Serum samples from patients with hepatitis and positive for hepatitis C virus subtype 1b (HCV-1b) and human pegivirus (HPgV, previously referred to as GBV-C) were obtained from the National Heart Lung and Blood Institute BioLINCC database (Mosley et al. 1990) (<https://biolincc.nhlbi.nih.gov/studies/>). Following pretreatment with nucleases, nucleic acid was extracted using the QIAamp Viral RNA Kit (Qiagen), and NGS libraries were prepared for Illumina sequencing using an Eppendorf epMotion 5075 BioRobot. For this study, a single NGS dataset from an individual positive for both HCV and HPgV infection and consisting of 47,920,144 100 bp PE reads was analyzed. The finding of HCV by NGS was independently confirmed using specific PCR followed by Sanger sequencing of the resulting 249 bp amplicon (Santos et al. 2012).

F. Hantavirus pulmonary syndrome (serum). A serum sample from a patient with hantavirus pulmonary syndrome during the 2012 Yosemite Valley outbreak (Nunez et al. 2014) was confirmed to be hantavirus positive by PCR using the following 2 primer sets: Hanta-1F (5'-AGAATGGTAGGGTGGGTTTGCAT-3') and Hanta-1R (5'-GGAGGAATATTACATGTGCCTTT-3'); Hanta-2F (5'-TATACAGAATGGTAGGGTGGGTTTGCAT-3') and Hanta-2R (5'-ATGTGTTTGATTACAGGCCAAATCATAA'-3). Following nucleic acid extraction, RNA was treated with Turbo DNase (Life Technologies, Carlsbad, CA), and then NGS libraries were prepared from both untreated and DNase-treated samples. The NGS run yielded a total of 28,635,304,250 bp PE reads, which were analyzed using SURPI, resulting in the identification of the Sin Nombre virus (SNV) strain Convict Creek 107 (Li et al. 1995; Schmaljohn et al. 1995).

G. Encephalitis (CSF). A clinical cerebrospinal fluid (CSF) sample was banked as part a study evaluating novel diagnostics for meningitis and encephalitis. Infection with human

herpesvirus 3 had been previously confirmed by PCR performed by a clinical reference laboratory. Nucleic acid was extracted using the EZ1 Viral Mini Kit (Qiagen, Carlsbad, CA), followed by NGS library construction using the Thruplex-FD Kit (Rubicon Genomics, Ann Arbor, Michigan). A total of 28,008,490 NGS reads obtained from a 250 bp PE Illumina MiSeq run were analyzed using SURPI.

I. Prostate biopsy (tissue). Nucleic acid extraction was performed and TruSeq-adapted NGS libraries were constructed as previously described (Lee et al. 2012). A total of 14,199,558 NGS reads obtained from a 150 bp PE Illumina MiSeq run were analyzed.

J. Prostate cancer (cell line). An NGS dataset derived from a prostate cancer primary cell line and consisting of 13,654,861 50 bp reads (corresponding to the first read only) was downloaded from the NIH Sequence Read Archive (SRA accession number SRR073726 (Prensner et al. 2011). This dataset harboring human papillomavirus type 18 (HPV-18) has been previously used as a benchmarking standard for the RINS (Bhaduri et al. 2012) and READSCAN (Naeem et al. 2013) algorithms.

K. Simian pneumonia (lung swab / tissue). An NGS dataset corresponding to an outbreak of fulminant pneumonia in a colony of New World titi monkeys caused by titi monkey adenovirus (TMAdV) (Chen et al. 2011) and consisting of 12,393,506 75 bp single-end reads was analyzed.

L. Spiked HIV (plasma). Recombinant human immunodeficiency virus type 1 (HIV-1) isolate DJ26 strain CRF02_AG (accession number AF063223) was cultured, purified as previously described (Michael et al. 1999), quantified using the Abbott RealTime HIV-1 assay (Abbott Molecular, Des Plaines, IL), and spiked at copy numbers of 10^2 , 10^3 , and 10^4 per mL into BaseMatrix defibrinated human plasma (Seracare) testing negative for hepatitis B virus, HCV, and HIV. Samples were pretreated with nucleases and nucleic acid was extracted using the EZ1 Viral Mini Kit (Qiagen, Carlsbad, CA). After library construction using the Thruplex-FD Kit (Rubicon Genomics, Ann Arbor, Michigan), a total of 11,734,074 NGS reads obtained from a 150 bp PE Illumina MiSeq run were analyzed.

M. Hemorrhagic fever outbreak, Democratic Republic of the Congo (DRC), Africa (serum). Serum samples from 48 cases of unexplained acute hemorrhagic fever were extracted using the QIAamp Viral RNA Kit (Qiagen), and TruSeq-adapted NGS libraries were generated using an epMotion 5075 BioRobot. An NGS dataset corresponding to a single individual with hemorrhagic fever from Katanga province in DRC and consisting of 10,632,074 100 bp PE reads was used for this analysis. The presence of *Plasmodium falciparum* DNA was

confirmed using specific PCR and Sanger sequencing of the 862 bp product as previously described (Hsiang et al. 2010).

N. Acute febrile illness, Tanzania, Africa (serum). Serum from 54 patients who died of undiagnosed critical febrile illness were pretreated with DNase, followed by nucleic acid extraction using the QIAamp UltraSens Virus Kit (Qiagen). After extraction, half of the extract was post-treated with DNase and then combined with the untreated half for the generation of individually barcoded NGS libraries. An NGS library corresponding to an individual positive for *Haemophilus influenzae* and containing 6,740,142 100 bp PE reads was analyzed. *H. influenzae* was confirmed by PCR to the *fucK* gene and Sanger sequencing of the 179 bp product as previously described (Abdeldaim et al. 2013).

Prospective clinical sample (serum). A serum sample from a returning traveler with an acute febrile illness was pretreated with DNase, followed by nucleic acid extraction using the EZ1 Viral Mini Kit (Qiagen, Carlsbad, CA) and NGS library preparation using a TruSeq-adapted protocol. A total of 10,196,620 reads from the Illumina MiSeq run were analyzed for pathogens using SURPI. From that run, 2,008,883 reads corresponded to acute serum from the returning traveler, while the remaining 8,187,737 reads corresponded to clinical samples from an unrelated patient with an unknown meningoencephalitis (Wilson et al. 2014). Following the detection of human herpesvirus 7 (HHV-7), the presence of HHV-7 was confirmed by specific PCR performed in the clinical microbiology laboratory. In contrast, IgG / IgM antibody titers for HHV-7 (Focus Diagnostics, Cypress, CA) were negative for acute infection.

(Figure 4 and Supplemental Table 3)

Acute respiratory illness from Influenza A (H1N1) pdm09 (nasal swab). An NGS dataset corresponding to nasopharyngeal swab samples from patients with laboratory-confirmed or suspected cases of 2009 H1N1 influenza virus infection (Greninger et al. 2010), was analyzed. A total of 7,000,000 65 bp reads from the NGS dataset were selected for SURPI analysis.

Hardware

For testing and benchmarking, we ran SURPI on several different types of hardware, all using the Ubuntu 12.04.3 LTS operating system. A 64-core AMD Opteron 6380 CPU (2.5 GHz) server with 512 GB random access memory (RAM) and 3 x 4 TB 7200 rpm hard drives was used for SURPI benchmarking. A secondary 32-core AMD Opteron 6282 CPU (1.4 GHz) server with 256 GB RAM and 3 x 4 TB 7200 rpm hard drives was used for routine SURPI processing.

For the cloud implementation of SURPI, a hi1.4xlarge (16v CPU, 60.5 GB, 2 x 1024 GB SSD) Amazon EC2 instance was used, with a dedicated EBS volume for persistent storage.

Reference databases

A 3.1 gigabase (Gb) human nucleotide database (human DB) was constructed from a combination of human genomic DNA (GRCh37 / hg19), rRNA (RefSeq), mRNA (RefSeq), and mitochondrial RNA (RefSeq) sequences in NCBI as of March of 2012. The bacterial nucleotide, viral nucleotide, and viral protein databases used by SURPI in fast mode (bacterial DB, viral nucleotide DB, and viral protein DB, respectively) were also constructed from sequences in NCBI as of March of 2012. The 3 Gb bacterial DB was constructed from all bacterial RefSeq entries and consisted of 348,922 unique accessioned sequences, each with a minimum length of 100 bp. The 1.4 Gb viral nucleotide DB included 1,193,607 entries and was constructed by searching for all viral sequences in the 42 Gb National Center for Biotechnology Information (NCBI) nt collection using the query term “viridae[Organism]” in BioPython (Cock et al. 2009). The viral protein DB was similarly constructed by extracting viral sequences from the NCBI nr DB collection. Index tables for SNAP (v0.15) were generated with an empirically determined default seed size of 20 for the human DB and viral nucleotide DB, and seed size of 16 for the bacterial DB. Index tables for RAPSearch (v2.09) (Zhao et al. 2012) were generated from the viral protein DB using default parameters.

To generate the National Center for Biotechnology Information (NCBI) nucleotide (nt) collection (NCBI nt DB) used by SURPI in comprehensive mode, the complete 42 Gb nucleotide collection (nt) was downloaded from NCBI in January of 2013. This collection consists of a comprehensive archive of sequences from multiple sources, including GenBank, European Molecular Biology Laboratory (EMBL), DNA Data Bank of Japan (DDBJ), and Protein Data Bank (PDB), and is the richest collection of annotated microbial sequence data publicly available. As SNAP uses 32-bit offsets in the reference genome during hashing, the aligner restricts the size of the reference genome to an absolute maximum of 2^{32} bases, or ~4.2 Gb. Thus, the 42 Gb NCBI nt collection was first split into 29 sub-databases, each approximately 1.5 Gb in size. Each sub-database was then indexed separately by SNAP at default parameters with a seed size of 20. This generated 29 SNAP indexed databases, each approximately 27 GB in size, with the aggregate of all 29 databases referred to as the NCBI nt DB.

***In silico* query datasets**

In silico query nucleotide datasets were generated using the read simulator wgsim (v.0.2.3) in the SAMtools distribution package (Li et al. 2009) with a 1% error model (Shendure

and Ji 2008), no paired ends, and fixed read length of 100 bp. The FASTA-formatted sequences inputted into wgsim were derived from the human, bacterial, and viral nucleotide DBs described above. The 1.25mil, 25mil, 125mil, and 1.25bil datasets generated by wgsim comprised a total of exactly 1,282,006, 25,032,000, 125,032,000 and 1,246,556,696 reads, respectively (Supplemental Table S2).

To determine the effect of read length on the accuracy of SNAP and RAPSearch classification (Supplemental Fig. S1), an *in silico* query dataset of ~1.25 million 800 base pair (bp) reads was generated using the read simulator wgsim with a 1% error model and no paired ends. The randomly generated input reads were derived from the human database (DB), bacterial DB, and viral nucleotide DBs described above. The 800 bp query dataset was then cropped to generate additional 650 bp, 500 bp, 350 bp, 250 bp, 150 bp, 100 bp, and 50 bp datasets.

A small query dataset was constructed by combining 25,000 randomly generated *in silico* 100 bp viral reads from the 1.25 million read dataset to 1000 random human reads from chromosomes 1, 2, 4, 5, 7, 9, 11, 13, 15, 17, 19, and 21. This 37k-100 read dataset was used to determine optimal alignment parameters for BWA and Bowtie2 (BT2) (Supplemental Fig. S4) as described below.

Range of parameters used for ROC curve analysis

Given the large number of combinations of parameters to test for BWA (Li and Durbin 2009) (ver 0.6.1-r104) and BT2 (Langmead and Salzberg 2012) (v2.0.0-beta7), varying 3 key parameters for BWA and Bowtie2 (BT2) across the full range was computationally prohibitive using the 1.25 million read query dataset. Thus, a selection of parameters was chosen for the receiver operating characteristic (ROC) curves in Figs. 2 and 4 by using the smaller 37k-100 dataset for initial ROC analysis testing 352 and 400 parameter combinations for BWA and BT2, respectively (Supplemental Fig. S4).

For BWA, the seed (-k) was varied across k2 (default), k8, k20, and k32, the maximum edit distance (-n) was varied across n0.04 (default), n0.1, n1, n2, n4, n8, n12, n16, n20, n24, and n28, and the seed length (-l) was varied across l2, l4, l8, l16, l24, l28, l32 (default), and l36. The best 23 parameter combinations as determined by calculating Youden's index and the F_1 score (Akobeng 2007) were used for plotting the BWA ROC curves (Figure 2A-E; Figure 4A-C), as follows: (k2, n0.04/n0.1/n8, l8/l16/l24/l28/l32/l36) and (k8, n0.04/n0.1/n4, l32), (k8, n12, l36), and (k2,n0.04,l4).

For BT2, the number of seed extensions (-D) was varied across D5, D15 (default), D40, and D100, the number of seed extension attempts (-L) was varied across L12, L20, L22 (default), L26, and L32, and the interval between seed substrings (-i) was varied across i0,0.01; i0,0.1; i0,0.5; i0,0.75; i0,1.15; i0,1.5; i0,1.75; i0,2.0; i0,2.5; i0,3.5; i1,0.01; i1,0.1; i1,0.5; i1,0.75; i1,1.15 (default); i1,1.5; i1,1.75; i1,2.0; i1,2.5; i1,3.5. The best 78 parameter combinations were used for plotting the points on the BT2 ROC curves (Figure 2A-E; Figure 4A-C), as follows: (D15/D10, L22/L26/L32, i0,0.75/i0,1.5/i0,1.75/i0,2.0/i0,2.5/i0,3.5/i1,1/i1,1.15/i1,1.5/i1,1.75/i1,2.0/i1,2.5/i1,3.5).

The expectation value (E-value) for BLASTn (NCBI BLAST 2.2.25 release) is the probability that a given alignment to a reference in the database was due purely to chance (Altschul et al. 1990). The following E-value cutoffs were used to generate the BLASTn ROC curves: 10^{-5} , 10^{-8} , 10^{-10} , 10^{-15} , 10^{-20} , 10^{-30} , 10^{-41} , 10^{-42} , 10^{-43} , 10^{-44} , 10^{-45} , 10^{-46} , 10^{-47} , 10^{-48} , 10^{-49} , 10^{-50} , with the default low-complexity filter turned off. For SNAP (v0.15), the ROC curves were generated by varying the threshold cutoffs for the edit distance parameter “d”, defined as the minimum number of substitutions, insertion, and deletions between a seed and candidate matched location in the reference genome, as follows: d2, d4, d6, d8, d10, d12, d14, d16, d18, d20, d22, d24, d26, d28.

For BLASTx (NCBI BLAST 2.2.25 release) and RAPSearch (v2.12), the following E-value cutoffs were used for the ROC curves: 10^1 , 10^0 , 10^{-1} , 10^{-3} , 10^{-5} , 10^{-8} , 10^{-10} , 10^{-15} , 10^{-20} , 10^{-30} .

Speed benchmarking for aligners

For speed benchmarking of nucleotide alignments to the human DB or translated nucleotide alignments to the viral protein DB (Figs. 3A and 3C), each algorithm used the maximum number of cores available to the server (n=64). Processing times were determined from the elapsed wall-clock time as measured by the computer’s system clock. The memory cache was cleared before each timing run. Optimal parameters for BWA and BT2 were obtained from the ROC plot of *in silico*-generated data (Fig. 2A). For BLASTn / BLASTx speed benchmarking, query read files were divided into 64 subfiles, and each subfile was run separately but in parallel using a single core. The time for each core to complete its processing was measured, and the average time per core was taken as the processing time. For read datasets in which the total computational time was prohibitive (Fig. 3A, asterisks), processing time was estimated from the processing time corresponding to smaller datasets and assuming a linearly proportional relationship between processing time and number of reads. The

processing time of SNAP alignment to the NCBI nt DB (Fig. 3B) was the total time taken to sequentially align reads to the 29 split nt sub-databases. After SNAP processing of each nt sub-database, a Python program was executed to scan through all new alignments and update the read classifications with the current “best” hits. We did not speed benchmark BWA, BT2, and BLASTn for alignments to the NCBI nt DB, as mapping to the 42 GB nt DB using these algorithms would be computationally prohibitive.

Optimization of *de novo* assembly

After SNAP alignment to the NCBI nt DB, reads that remained unclassified were combined with reads identified as viral, and duplicate reads were removed. The combined dataset was then *de novo* assembled prior to translated nucleotide (protein) alignment. Benchmarking of *de novo* assembly approaches was performed using NGS clinical sample datasets containing sequences from target viral pathogens (Fig. 5, Supplemental Table S4).

Three methods of *de novo* assembly were compared for each dataset:

- (1) SOAPdenovo (v2.04 release) (Li et al. 2010) was used with the default configuration file (“SOAPdenovo-63mer all -s config_file -K”), while varying kmer length (-K) as follows: 15, 25, 35, 45, 55, 65, and 75. As each dataset required a different kmer parameter for optimal assembly, a kmer of 35 was selected as a representative parameter closest to the optimal AbySS kmer length of 34.
- (2) abyss-pe, the MPI (message passing interface)-based parallel version of the AbySS *de novo* assembler (v1.3.5 release) (Simpson et al. 2009), was run at default parameters using the empirically determined optimal kmer length of 34.
- (3) AbySS was run multiple times in succession, using both the entire dataset and individually partitioned sets of 100,000 reads as input. Output contig sequences of length greater than or equal to the read length were then assembled using the Overlap-Layout-Consensus (OLC) *de novo* assembler Minimo (v1.6 release) (Treangen et al. 2011) at default parameters.

After *de novo* assembly, contigs generated with length greater than 1.75x the read length were retained. Output contigs were then mapped to the target viral genome, and the number of mapped contigs, percent coverage, and corresponding N50 statistic were calculated (Supplemental Table S22). Method 3 was found to generate the longest contigs, and was thus incorporated into the SURPI pipeline

Running the SURPI pipeline

The scripts, external software dependencies, and database dependencies for the SURPI pipeline are shown in Supplemental Fig. S5. **Software dependencies:** SURPI currently runs on Ubuntu 12.04, and requires the following open-source software: Perl, Python, cutadapt (Martin 2011), prinseq (Schmieder and Edwards 2011), gt sequniq (Gremme et al. 2013), SNAP (Zaharia et al. 2012), fqextract (<http://www.biostars.org/p/10353/>, last accessed November 20th, 2013), ABySS (Simpson et al. 2009), Minimo (Treangen et al. 2011), RAPSearch2 (Zhao et al. 2012), and seqtk (<https://github.com/lh3/seqtk/>, last accessed November 20th, 2013). **Database dependencies:** Currently, SNAP uses the human DB for computational subtraction. In fast mode, SURPI uses the curated bacterial and viral nucleotide DBs, while in comprehensive mode, SNAP uses the NCBI nt DB. RAPSearch uses a viral protein database derived from NCBI nr DB. To obtain taxonomic annotation associated with each assigned sequence data identifier (GI or accession number), an SQLite taxonomy database was generated from the NCBI taxonomy databases available at <ftp://ftp.ncbi.nih.gov/pub/taxonomy/>. All SNAP and RAPSearch databases are updated at least twice a year.

SURPI execution: SURPI can be run as a single Linux script, with a configuration file allowing the user to set the following parameters:

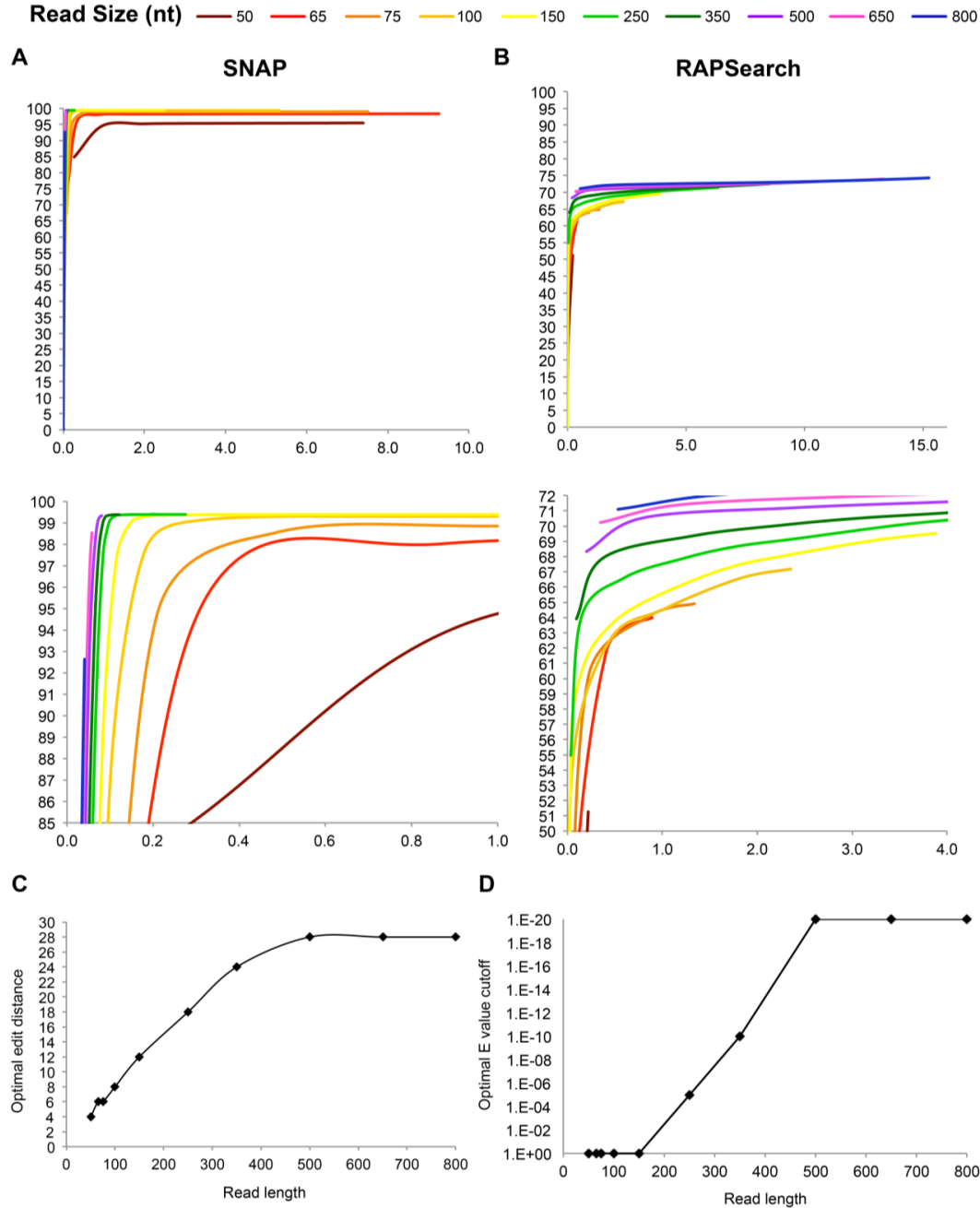
- Input FASTQ file
- Number of cores to engage for parallelized processing steps (default = all available cores)
- Sanger or Illumina quality cutoffs (default = Sanger)
- Smallest read length allowed after quality filtering (default = 50 bp)
- Parameters for read cropping prior to SNAP alignment (default = 10-85 bp)
- TruSeq or Nextera adapters for Illumina libraries (default = TruSeq)
- Location of the SNAP indexed databases
- Location of the RAPSearch indexed databases
- SNAP edit distance (default = 12)
- RAPSearch E-value cutoff (default = 1)
- Run translated nucleotide alignment using RAPSearch to the viral protein database or the entire NR protein database (default = viral protein database)

Generation of SURPI summary tables

Summary tables are outputted after alignment to the bacterial DB and viral nucleotide DBs in fast mode and after alignment to the NCBI nt DB and viral protein DB in comprehensive mode. For each of the following 4 taxonomic categories (GenBank gi identifier, species, genus, and family), classified reads are tabulated into an individual summary table. For multiplexed NGS data, each indexed barcode in the summary table is displayed as a separate column. Examples of summary table output for viral species identification at the nucleotide level (Tables S6-13) and at the amino acid level (Table S14-21) are provided.

GenBank accession numbers and abbreviations

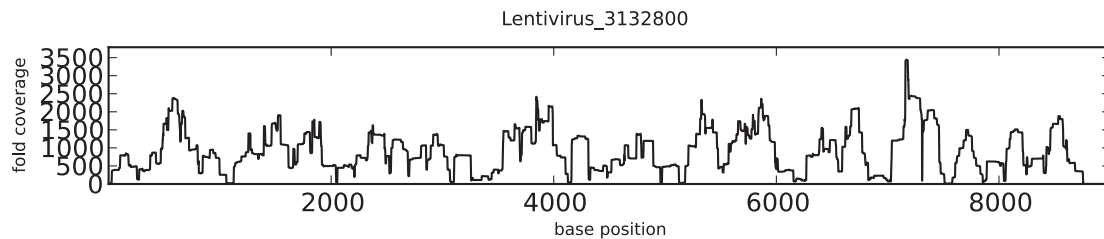
BASV, Bas-Congo virus (JX297815); **bat influenza H17N10**, influenza A virus A/little yellow-shouldered bat/Guatemala/2009/ (H17N10) (CY103873 - CY103896); **SNV**, Sin Nombre virus (HPSCC107MS, HPSCC107LS, and HPSCC107SS); **H. influenzae**, *Haemophilus influenzae* PittEE (CP000671); **HCV-1b**, hepatitis C virus subtype 1b (EU256064); **HHV-3**, human herpesvirus 3 (JF306641); **HHV-7**, human herpesvirus 7 strain RK (AF037218.1); **HIV-1**, human immunodeficiency virus 1 isolate DJ263, subtype CRF02_Ag (AF063223); **HPeV-1**, human parechovirus 1 (GQ183034); **HPV-18**, human papillomavirus type 18 isolate CU16 (GQ180792); **influenza A (H1N1)pdm2009**, influenza A virus A/California/04/2009 (H1N1) (JF915184 - JG15191); **norovirus** (EU310927); **P. falciparum**, *Plasmodium falciparum* clone 3D7 (NC004325, NC_000910, NC_000521, NC_004318, NC004326, NC_004327, NC_004328, NC_004329, NC_004330, NC_004314, NC_004315, NC_004316, NC_004311, AL844509, AE014187); **RSV**, respiratory syncytial virus (JN032116); **SAdV-18**, simian adenovirus 18 (FJ025931); **SaV**, sapovirus NongKhai-50/Thailand (AY646853); **TMAAdV**, titi monkey adenovirus (HQ913600); **ZEBOV**, Zaire ebolavirus (AY354458).



Supplemental Figure S1. Effect of read length on accuracy of SNAP and RAPSearch classification. Receiver operating characteristic (ROC) curve analysis of (A) SNAP performance and (B) RAPSearch performance in classifying viral reads at read lengths of 50, 65, 75, 100, 150, 250, 350, 500, 650, and 800 bp. Sensitivity or TPR (y-axis) of detection is plotted against 1-specificity or FPR (x-axis). The lower panels are plotted on a different scale to show separation between the curves. (C) Plot of optimal edit distance against read length following SNAP alignment. (D) Plot of optimal E-value cutoff against read length following RAPSearch alignment. Optimal parameters were determined by Youden's Index and F_1 score.

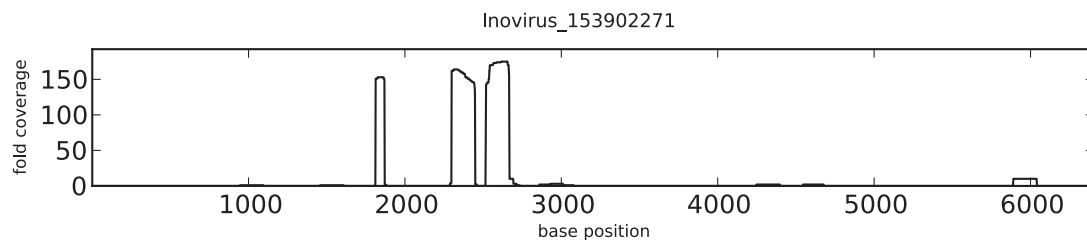
Supplemental Figure S2. Coverage maps automatically generated by SURPI in comprehensive mode. For each clinical NGS dataset, coverage maps of top 3 viral genera and the closest reference genome for each genus (ranked by percent genome coverage) are displayed. (A) spiked HIV plasma (10^4 copies / mL). (B) pediatric diarrhea (stool). (C) prostate cancer (cell line). (D) encephalitis (CSF). (E) TTVS hepatitis (serum). Coverage maps with 3 or fewer viral reads are not shown. The read coverage (y-axis, “fold coverage”) is plotted as a function of nucleotide position along the genome (x-axis, “base position”).

A



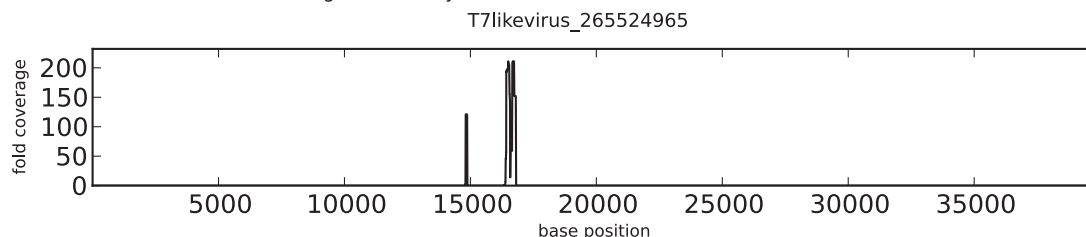
mapping Lentivirus.bar.#GCCAAT@_HIV.NT.snap.matched.Viruses
against Lentivirus.3132800.fasta with gi definition
gi|3132800|gb|AF063223.1| HIV-1 isolate DJ263 from Djibouti, complete genome

Reference sequence length = 9002 bp
Coverage in bp = 8970
%Coverage = 99.644523
Average depth of coverage = 892.990335
Number of reads contributing to assembly = 54537



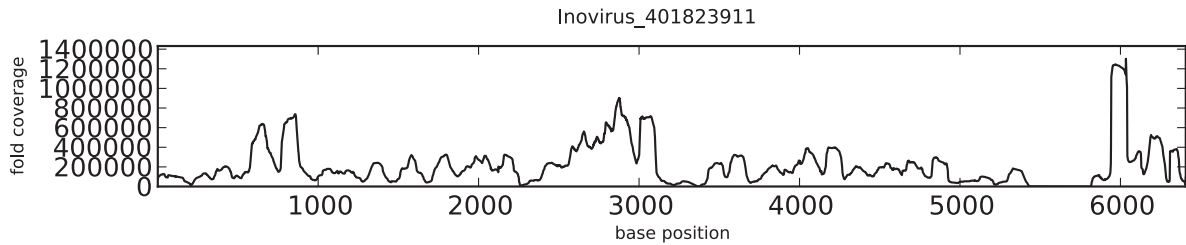
mapping Inovirus.bar.#GCCAAT@_HIV.NT.snap.matched.Viruses
against Inovirus.153902271.fasta with gi definition
gi|153902271|dbj|AB334721.1| Enterobacteria phage f1 DNA, complete genome, isolate: NBRC 20015

Reference sequence length = 6407 bp
Coverage in bp = 1422
%Coverage = 22.194474
Average depth of coverage = 9.526611
Number of reads contributing to assembly = 358



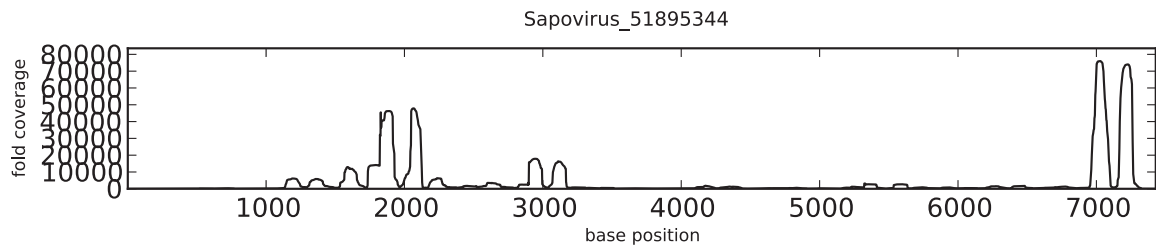
mapping T7likevirus.bar.#GCCAAT@_HIV.NT.snap.matched.Viruses
against T7likevirus.265524965.fasta with gi definition
gi|265524965|gb|GU071091.1| Enterobacteria phage T7, complete genome

Reference sequence length = 39778 bp
Coverage in bp = 563
%Coverage = 1.415355
Average depth of coverage = 1.792171
Number of reads contributing to assembly = 544

B

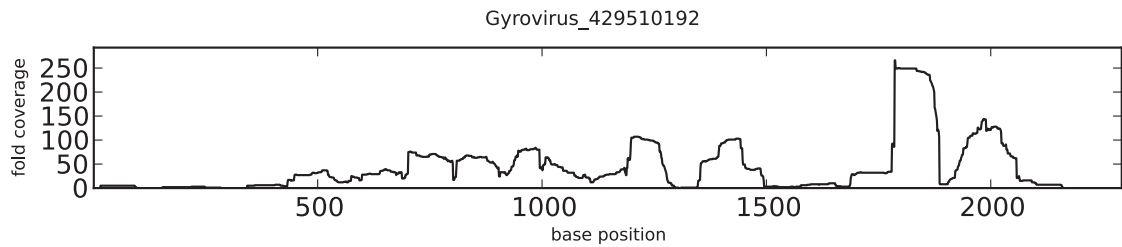
mapping Inovirus.bar.#0@_Mexus.NT.snap.matched.Viruses
 against Inovirus.401823911.fasta with gi definition
 gi|401823911|gb|JX412914.1| Enterobacteria phage M13 isolate WT variety Rutgers, complete genome

Reference sequence length = 6407 bp
 Coverage in bp = 6357
 %Coverage = 99.219603
 Average depth of coverage = 213920.240362
 Number of reads contributing to assembly = 14778987



mapping Sapovirus.bar.#0@_Mexus.NT.snap.matched.Viruses
 against Sapovirus.51895344.fasta with gi definition
 gi|51895344|gb|AY694184.2| Sapovirus Hu/Dresden/pJG-Sap01/DE, complete genome

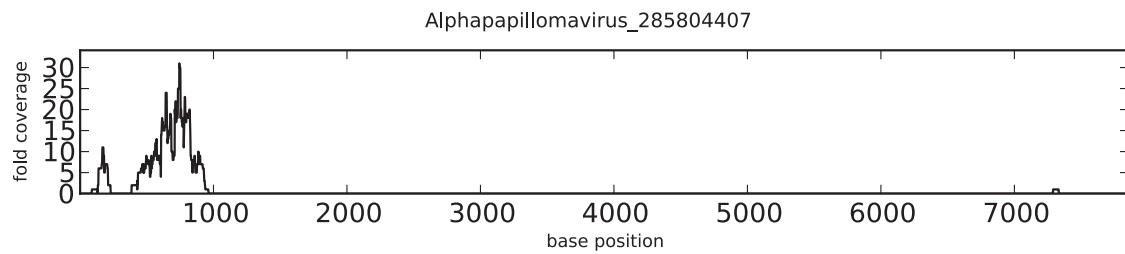
Reference sequence length = 7429 bp
 Coverage in bp = 7319
 %Coverage = 98.519316
 Average depth of coverage = 4469.658231
 Number of reads contributing to assembly = 371633



mapping Gyrovirus.bar.#0@_Mexus.NT.snap.matched.Viruses
 against Gyrovirus.429510192.fasta with gi definition
 gi|429510192|gb|JX964755.1| Chicken anemia virus isolate GXC060821, complete genome

Reference sequence length = 2292 bp
 Coverage in bp = 2027
 %Coverage = 88.438045
 Average depth of coverage = 40.489965
 Number of reads contributing to assembly = 1080

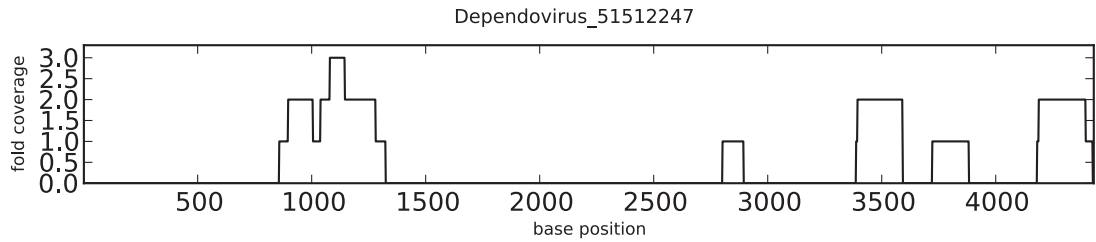
C



mapping Alphapapillomavirus.bar.#80@_SRR073726.R1.50.NT.snap.matched.Viruses
against Alphapapillomavirus.285804407.fasta with gi definition
gi|285804407|gb|GQ180792.1| Human papillomavirus type 18 isolate CU16, complete genome

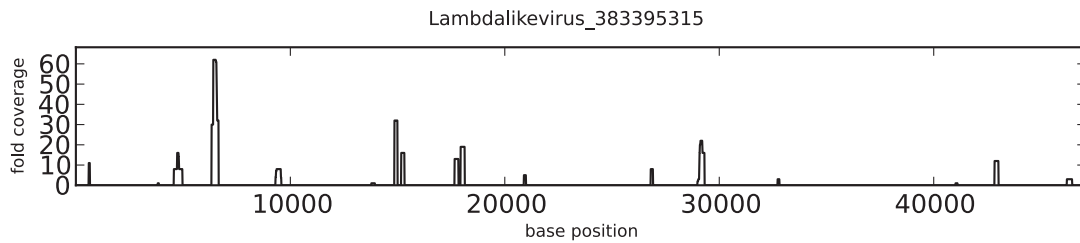
Reference sequence length = 7857 bp
Coverage in bp = 758
%Coverage = 9.647448
Average depth of coverage = .841288
Number of reads contributing to assembly = 152

D



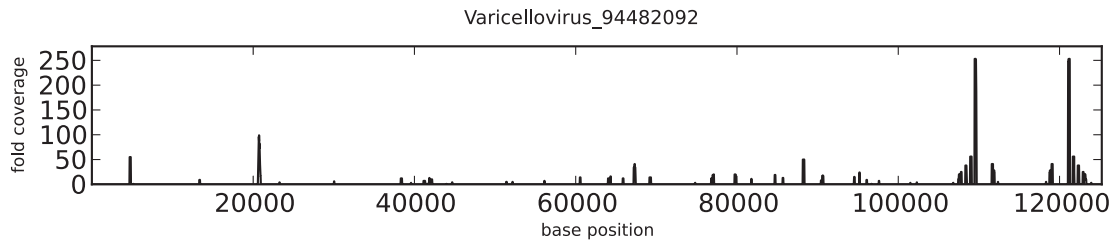
mapping Dependovirus.bar.#0@_VZV.NT.snap.matched.Viruses
against Dependovirus.51512247.fasta with gi definition
gi|51512247|gb|AY695376.1| Adeno-associated virus isolate hu.S17 Rep78 protein and capsid protein...

Reference sequence length = 4430 bp
Coverage in bp = 1170
%Coverage = 26.410835
Average depth of coverage = .449435
Number of reads contributing to assembly = 10



mapping Lambdalikeyvirus.bar.#0@_VZV.NT.snap.matched.Viruses
against Lambdalikeyvirus.383395315.fasta with gi definition
gi|383395315|gb|JQ086376.1| Enterobacteria phage HK630, complete genome

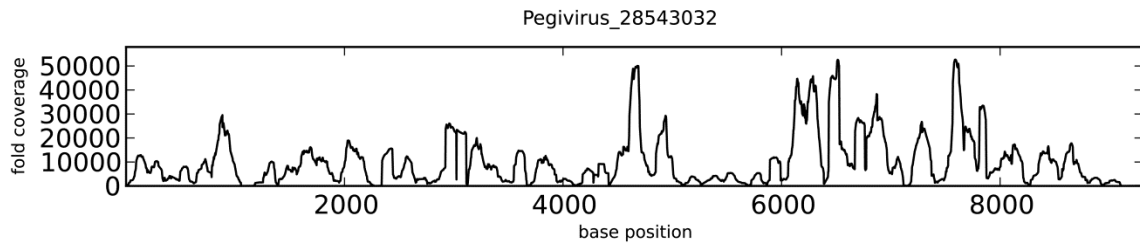
Reference sequence length = 47090 bp
Coverage in bp = 3142
%Coverage = 6.672329
Average depth of coverage = .951306
Number of reads contributing to assembly = 233



mapping Varicellovirus.bar.#0@_VZV.NT.snap.matched.Viruses
against Varicellovirus.94482092.fasta with gi definition
gi|94482092|gb|DQ479957.1| Human herpesvirus 3 strain 03-500, complete genome

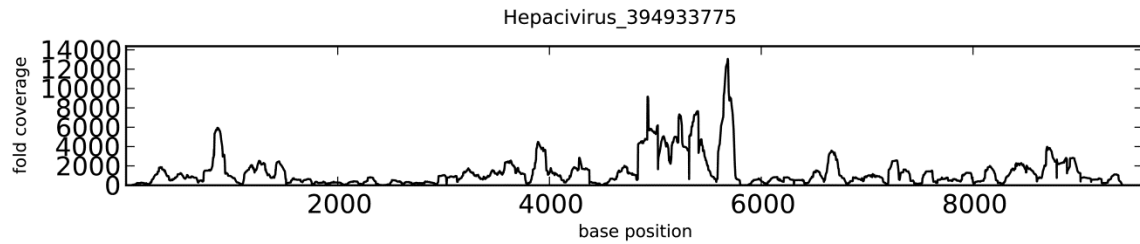
Reference sequence length = 125239 bp
Coverage in bp = 7977
%Coverage = 6.369421
Average depth of coverage = 2.241378
Number of reads contributing to assembly = 985

E



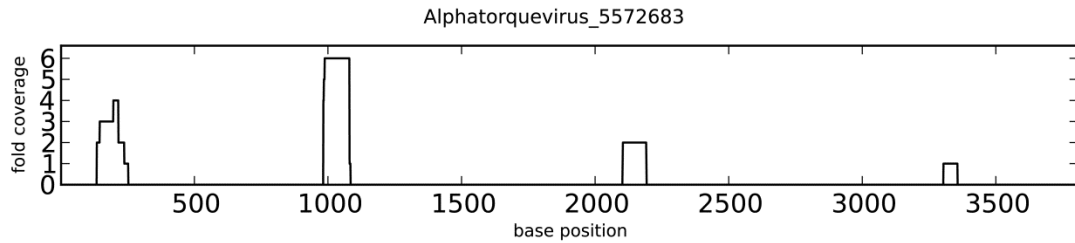
mapping Pegivirus.bar#CACGAT.TTVS.NT.snap.matched.Viruses
against Pegivirus.28543032.fasta with gi definition
gi|28543032|gb|AY196904.1| Hepatitis G virus isolate 765, complete genome

Reference sequence length = 9309 bp
Coverage in bp = 9251
%Coverage = 99.376947
Average depth of coverage = 10126.536684
Number of reads contributing to assembly = 1081985



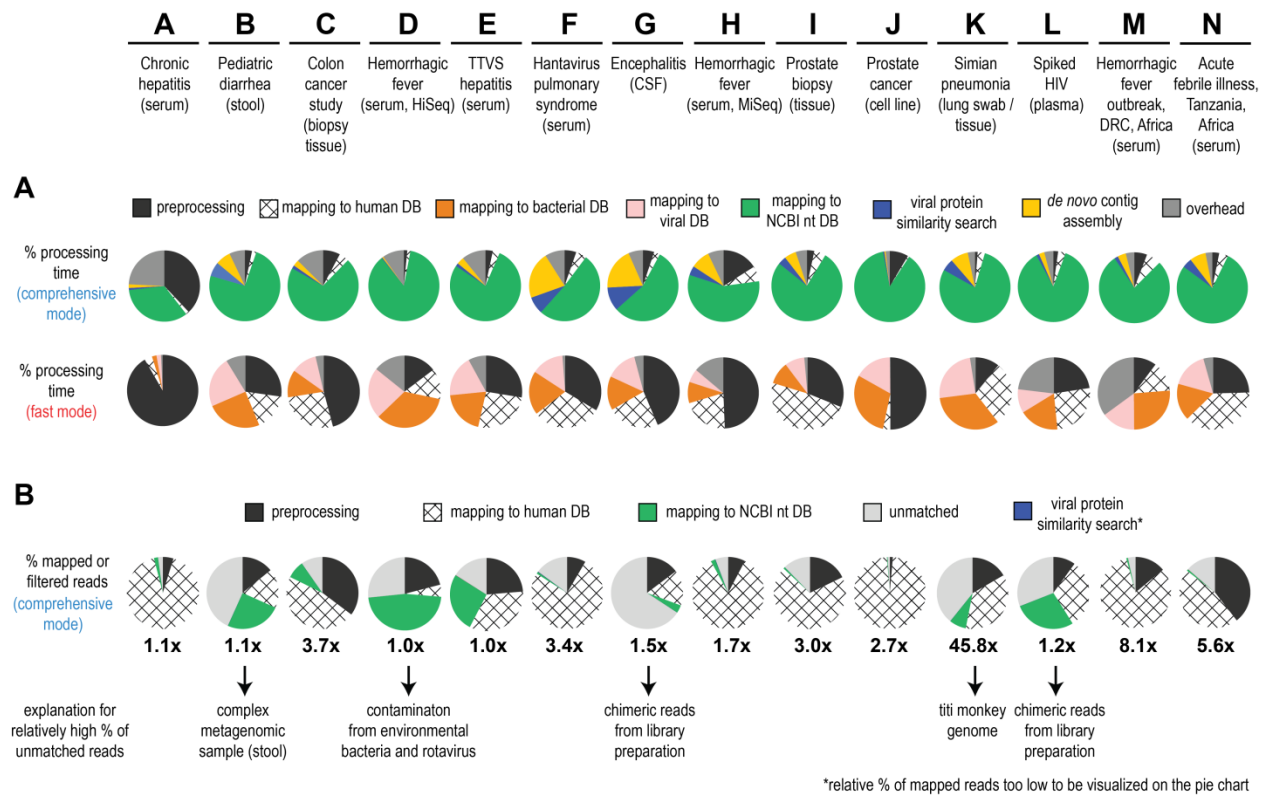
mapping Hepacivirus.bar#CACGAT.TTVS.NT.snap.matched.Viruses
against Hepacivirus.394933775.fasta with gi definition
gi|394933775|dbj|AB691953.1| Hepatitis C virus subtype 1b genomic RNA, complete genome, strain: NC1

Reference sequence length = 9607 bp
Coverage in bp = 9384
%Coverage = 97.678775
Average depth of coverage = 1363.320703
Number of reads contributing to assembly = 146259

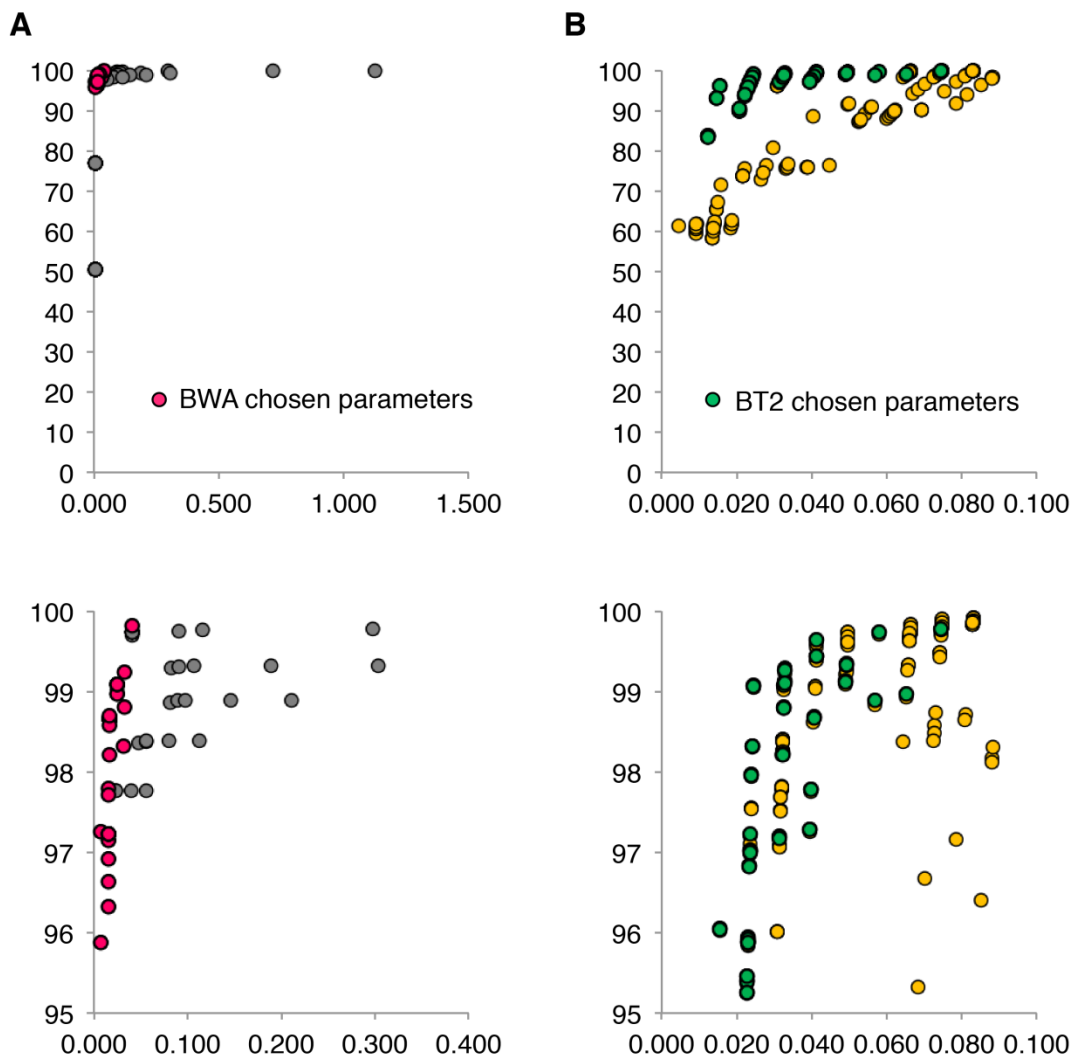


mapping Alphatorquevirus.bar.#CACGAT@_.TTVS.NT.snap.matched.Viruses
against Alphatorquevirus.5572683.fasta with gi definition
gi|5572683|dbj|AB025946.2| Torque teno virus 19 DNA, complete genome, isolate: TTV SANBAN

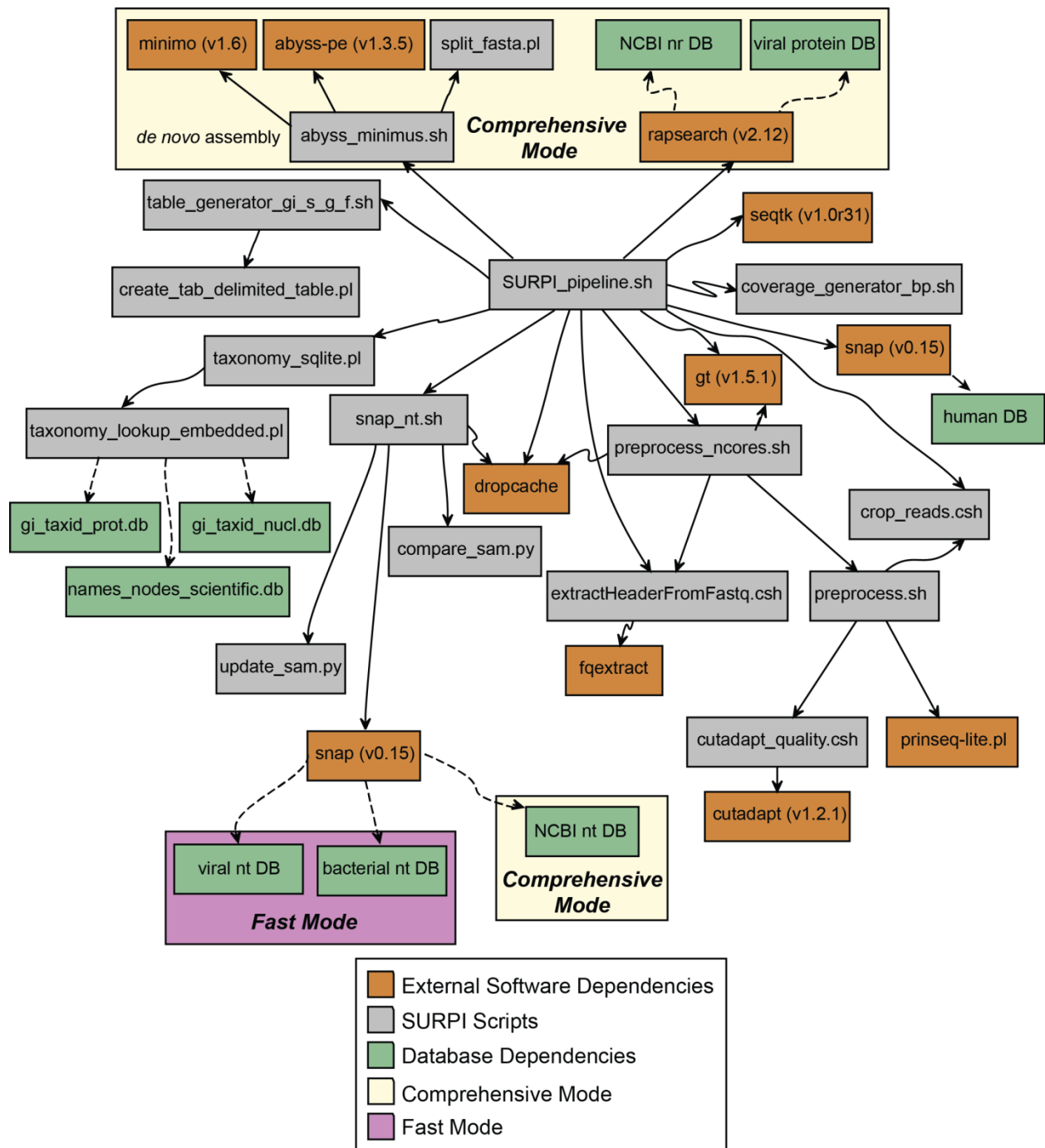
Reference sequence length = 3808 bp
Coverage in bp = 363
%Coverage = 9.532563
Average depth of coverage = .295693
Number of reads contributing to assembly = 13



Supplemental Figure S3. Relative percentages of processing time and mapped / filtered reads for clinical metagenomic datasets analyzed using SURPI. (A) Pie charts representing the percentage of total processing time corresponding to each SURPI step in comprehensive and fast modes. **(B)** Pie charts representing the percentage of preprocessed reads and reads aligned at each SURPI step in comprehensive mode. The number below the pie chart shows the fold-increase in the number of reads classified in comprehensive mode relative to fast mode. Arrows point to likely explanations for the relatively high percentage of unmatched reads in some samples.



Supplemental Figure S4. ROC curves generated from varying BWA and BT2 parameter combinations. ROC curves were generated to evaluate the ability of BWA and BT2 to correctly classify a query dataset of 37k 100 bp *in silico*-generated reads when mapped against the viral nucleotide DB. The sensitivity or TPR (y-axis) is plotted against 1-specificity or FPR (x-axis). (A) BWA ROC plots displaying 325 combinations of 3 parameters. Points in pink on the ROC plot refer to parameters retained for subsequent ROC curve analysis (Figs. 2 and 4). (B) BT2 ROC plots displaying 400 combinations of 3 parameters. Points in green refer to parameters retained for subsequent ROC curve analysis. The lower panels are magnifications of the upper panels.



Supplemental Figure S5. SURPI scripts and software / database dependencies

SUPPLEMENTAL TABLES

Supplemental Table S1. Processing time and cost of SURPI versus PathSeq v1.2 on a cloud server. SURPI in fast and comprehensive mode, as well as PathSeq, were run on their respective cloud-based AWS (Amazon Web Services) instances.

Dataset	Number of reads	SURPI (fast mode)		SURPI (comprehensive mode)		PathSeq	
		Time	Cost	Time	Cost	Time	Cost
Prostate cancer (cell line)	13,654,861	0:19:47	\$3.10	1:03:22	\$6.20	18:50:00	\$91.60
TTVS hepatitis (serum)	47,920,144	0:57:09	\$3.10	4:28:55	\$15.50	27:25:04	\$134.24
Pediatric diarrhea (stool)	158,026,920	2:50:17	\$9.30	14:43:12	\$43.40	~9.2 days	\$1,082.42

Supplemental Table S2. Composition and number of reads corresponding to *in silico*-generated NGS datasets. The number of reads derived from each reference sequence is tabulated. N/A, not applicable.

Origin	1.25mil	25mil	125mil	1.25bil	1mil-800	37k-100
chromosome 01	79,660	1,593,193	7,965,948	79,659,454	76,034	1,000
chromosome 02	77,522	1,550,442	7,752,211	77,522,114	80,481	1,000
chromosome 03	63,122	1,262,431	6,312,153	63,121,534	65,721	N/A
chromosome 04	61,241	1,224,840	6,124,197	61,241,978	63,375	1,000
chromosome 05	57,668	1,153,369	5,766,846	57,668,460	60,017	1,000
chromosome 06	65,092	1,301,816	6,509,080	65,090,802	56,519	N/A
chromosome 07	50,785	1,015,705	5,078,526	50,785,263	52,506	1,000
chromosome 08	46,679	933,583	4,667,917	46,679,179	48,199	N/A
chromosome 09	45,168	903,341	4,516,713	45,167,131	40,559	1,000
chromosome 10	43,203	864,060	4,320,299	43,202,990	44,249	N/A
chromosome 11	43,048	860,948	4,304,739	43,047,394	44,287	1,000
chromosome 12	42,667	853,331	4,266,657	42,666,565	44,063	N/A
chromosome 13	36,711	734,230	3,671,149	36,711,494	32,163	1,000
chromosome 14	34,219	684,374	3,421,869	34,218,687	29,829	N/A
chromosome 15	32,683	653,657	3,268,286	32,682,857	27,616	1,000
chromosome 16	28,801	576,029	2,880,144	28,801,437	26,624	N/A
chromosome 17	26,525	530,482	2,652,412	26,524,118	26,257	1,000
chromosome 18	24,889	497,784	2,488,923	24,889,225	25,181	N/A
chromosome 19	18,929	378,565	1,892,822	18,928,209	18,863	1,000
chromosome 20	20,090	401,800	2,008,998	20,089,984	20,087	N/A
chromosome 21	15,351	307,013	1,535,068	15,350,686	6,298	1,000
chromosome 22	16,354	327,076	1,635,382	16,353,819	11,748	N/A
mitochondrial	5	106	528	5,282	N/A	N/A
multiple chromosomal locations	1,174	23,428	117,147	1,171,486	N/A	N/A
human mRNA	N/A	N/A	N/A	N/A	39,735	N/A
X chromosome	49,494	989,879	4,949,397	49,493,968	51,024	N/A
Y chromosome	18,926	378,518	1,892,589	18,925,889	8,565	N/A
Zaire ebolavirus	1,000	1,000	1,000	1,000	1,000	N/A
HIV-1 CRF02_AG	1,000	1,000	1,000	1,000	1,000	N/A
BASV	1,000	1,000	1,000	1,000	1,000	N/A
norovirus	1,000	1,000	1,000	1,000	1,000	N/A
TMAv	1,000	1,000	1,000	1,000	1,000	N/A
influenza A (H1N1)pdm2009 virus	1,000	1,000	1,000	1,000	1,000	N/A
bat influenza A (H17N10)	1,000	1,000	1,000	1,000	1,000	N/A
bacterial DB	250,000	5,000,000	25,000,000	246,526,691	250,000	N/A
viral nt DB	25,000	25,000	25,000	25,000	25,000	25,000
Total number of reads	1,282,006	25,032,000	125,032,000	1,246,558,696	1,282,000	37,000

Supplemental Table S3. Clinical NGS datasets used for ROC curve analysis. The percent coverage and overall nucleotide identity shown are with respect to the closest viral genome in the reference database. Abbreviations: bp, base pairs; PCR, polymerase chain reaction.

	Pediatric diarrhea	Acute respiratory illness	Hantavirus pulmonary syndrome
sample type	stool	respiratory	Serum
target viral species	respiratory syncytial virus	influenza A (H1N1)pdm2009	Sin Nombre hantavirus
confirmatory testing	PCR	clinical PCR (Greninger et al. 2010)	PCR, serology (Nunez, et al., 2014)
challenge	complex metagenomic sample; virus with ~10% sequence divergence	short reads (65 bp)	longer reads (250 bp, with chimeric reads)
paired-end length	100 bp	65 bp	250 bp
# of raw reads	7,000,000	4,560,147	7,000,000
# reads after computational subtraction	5,002,838	773,835	1,944,847
# of true viral reads	1,780	53,749	611
% true viral reads	0.025	1.179	0.009
% coverage (% nucleotide identity)	19.6 (90.6)	92.0 (98.6)	91.6 (96.8)

Supplemental Table S4. Clinical NGS datasets used for SURPI performance validation.

Abbreviations: R1, read 1; R2; read 2; PCR, polymerase chain, bp, base pairs; SaV, sapovirus; HPeV-1, human parechovirus type 1; BASV, Bas-Congo virus; HCV-1b, hepatitis C virus, type 1b; HHV-3, human herpesvirus type 3; HPV-18, human papillomavirus virus type 18; TMAdV, titi monkey adenovirus; HIV-1, human immunodeficiency virus 1.

Dataset	Target pathogen*	Independent confirmatory testing	% coverage (% nucleotide identity)	# of raw reads (R1+R2) [‡]	# (%) of true microbial reads [‡]	Paired-end read length
A. chronic hepatitis (serum)	—	—	—	508,729,590	—	100 bp
B. pediatric diarrhea (stool)	SaV	PCR	99.10 (100)	158,026,920	388,404 (0.25)	75 bp
B. pediatric diarrhea (stool)	HPeV-1	PCR	94 (79.8)	158,026,920	122,850 (0.078)	75 bp
C. colon cancer study (biopsy tissue)	<i>F. nucleatum</i>	(Castellarin, et al., 2012)	N/A	143,419,974	144,650 (0.10)	85 bp
D. hemorrhagic fever (HiSeq,serum)	BASV	PCR, serology (Grard, et al., 2012)	2 (72)	100,000,000	17,937 (0.018)	100 bp
E. TTVS hepatitis (serum)	HCV-1b	PCR	97.9 (98.6)	47,920,144	224,554 (0.47)	100 bp
F. hantavirus pulmonary syndrome (serum)	—	—	—	28,635,304	—	250 bp
G. encephalitis (CSF)	HHV-3	clinical PCR	91.10 (100)	28,008,490	1,131 (0.004)	250 bp
H. hemorrhagic fever (MiSeq,serum)	BASV	PCR, serology (Grard, et al., 2012)	2 (72)	17,881,580	60,733 (0.34)	400 / 100 bp
I. prostate biopsy (tissue)	—	—	—	14,199,558	—	150 bp
J. prostate cancer (cell line)	HPV-18	(Prensner et al. 2011)	100 (100)	13,654,861	13,666 (0.10)	80 bp
K. simian pneumonia (lung swab)	TMAdV	PCR, serology (Chen et al. 2011)	40 (82)	24,787,012	13,289 (0.054)	65 bp
L. spiked HIV (plasma) 10 ² copies/mL	HIV-1	spike-in	99.40 (100)	2,091,530	56 (0.003)	150 bp
L. spiked HIV (plasma) 10 ³ copies/mL	HIV-1	spike-in	99.40 (100)	2,126,578	3,844 (0.18)	150 bp
L. spiked HIV (plasma) 10 ⁴ copies/mL	HIV-1	spike-in	99.40 (100)	2,522,034	55,705 (2.2)	150 bp
M. hemorrhagic fever outbreak, DRC (serum)	<i>P. falciparum</i>	PCR	N/A	11,634,074	2,880 (0.025)	100 bp
N. acute febrile illness, Tanzania (serum)	<i>H. influenzae</i>	PCR	N/A	10,632,192	117 (0.001)	100 bp

*as shown in Figs. 5 and 6

[‡]corresponding to the target pathogen

Supplemental Table S5. Sensitivity and specificity of viral detection in clinical NGS

datasets. Abbreviations: FAST, SURPI in fast mode; COMP, SURPI in comprehensive mode; RAP, RAPSearch translated nucleotide aligner, SaV, sapovirus; HPeV-1, human parechovirus type 1; BASV, Bas-Congo virus; HCV-1b, hepatitis C virus, type 1b; HHV-3, human herpesvirus type 3; HPV-18, human papillomavirus virus type 18; TMAAdV, titi monkey adenovirus; HIV-1, human immunodeficiency virus 1.

Datasets [detected virus]	Percent Sensitivity			Percent Specificity		
	FAST (SNAP)	COMP (SNAP)	COMP (RAP)	FAST (SNAP)	COMP (SNAP)	COMP (RAP)
B. Pediatric diarrhea (stool) [SaV]	83.99	97.37	76.17	100.00	100.00	100.00
B. Pediatric diarrhea (stool) [HPeV-1]	72.95	76.98	97.94	99.97	99.97	99.95
D. Hemorrhagic fever (Hiseq serum) [BASV]	0.00	0.00	15.75	N/A	N/A	100
E. TTVS hepatitis (serum) [HCV-1b]	95.98	96.30	78.85	99.98	99.98	99.94
G. encephalitis (CSF) [HHV-3]	89.66	89.66	29.17	100	100	100
H. hemorrhagic fever (MiSeq, serum) [BASV]	0.00	0.00	23.69	N/A	N/A	100
J. prostate cancer (cell line) [HPV-18]	98.60	93.37	25.83	100	100	100
K. simian pneumonia (lung swab) [TMAAdV]	7.95	8.04	13.79	100	100	100
L. spiked HIV (plasma) 10 ² copies/mL [HIV-1]	91.07	60.00	91.07	100	100	100
L. spiked HIV (plasma) 10 ³ copies/mL [HIV-1]	99.56	75.00	99.12	100	100	100
L. spiked HIV (plasma) 10 ⁴ copies/mL [HIV-1]	95.30	73.26	97.40	100	100	99.99

Supplemental Tables S6-S13. Viral species summary tables of SURPI (SNAP) alignment to the NCBI nt DB. Letters or numbers following the hash mark denote indexed barcodes for multiplexing. Abbreviations: HIV-1, human immunodeficiency type 1; SaV, sapovirus; HPeV-1, human parechovirus type 1; BASV, Bas-Congo virus; HCV-1b, hepatitis C virus, type 1b; HHV-3, human herpesvirus type 3; HPV-18, human papillomavirus virus type 18; TMAAdV, titi monkey adenovirus.

Supplemental Table S6. Viral summary table of SURPI alignment to NCBI nt DB for a spiked HIV NGS dataset (L), containing HIV-1.

Species	Genus	Family	Index #CGATGT (10 ² HIV)	Index #GCCAAT (10 ⁴ HIV)	Index #TGACCA (10 ³ HIV)
Acinetobacter phage IME-AB2			0	2	0
Acinetobacter bacteriophage AP22			0	270	0
Enterobacteria phage HK542			10	12	0
Pseudomonas phage PA7			418	0	0
Torque teno mini virus 5	Betatorquevirus	Anelloviridae	122	0	0
Torque teno mini virus 9	Betatorquevirus	Anelloviridae	124	0	0
SEN virus		Anelloviridae	158	0	8
Torque teno mini virus 7	Betatorquevirus	Anelloviridae	179	0	0
Torque teno mini virus 2	Betatorquevirus	Anelloviridae	2	0	0
Torque teno virus		Anelloviridae	2358	122	681
Micro Torque teno virus		Anelloviridae	3	0	0
Torque teno virus 10	Alphatorquevirus	Anelloviridae	3	0	0
TTV-like mini virus		Anelloviridae	49	0	170
Torque teno virus 11	Alphatorquevirus	Anelloviridae	8	0	0
uncultured Anellovirus		Anelloviridae	8	0	0
GB virus C	Pegivirus	Flaviviridae	14922	56027	55017
Human herpesvirus 1	Simplexvirus	Herpesviridae	0	10	0
Trichoderma hypovirus	Hypovirus	Hypoviridae	25	0	0
Enterobacteria phage f1	Inovirus	Inoviridae	3	0	0
Enterobacteria phage M13	Inovirus	Inoviridae	861	359	654
Megavirus courdo7		Mimiviridae	0	0	1
Acanthamoeba castellanii mamavirus		Mimiviridae	0	0	3
Acidithiobacillus phage AcaML1		Myoviridae	0	30	0
Pseudomonas phage EL	Phikzlikevirus	Myoviridae	400	49	14
Enterobacteria phage T3	T7likevirus	Podoviridae	0	0	1
Enterobacteria phage 13a	T7likevirus	Podoviridae	1	0	0
Enterobacteria phage T7M	T7likevirus	Podoviridae	1	0	0
Yersinia phage phiA1122	T7likevirus	Podoviridae	150	121	2093
Salmonella phage Vi06	T7likevirus	Podoviridae	40	0	83
Enterobacteria phage T7	T7likevirus	Podoviridae	535	423	1253
Aids-associated retrovirus		Retroviridae	0	0	18
Simian-Human immunodeficiency virus	Lentivirus	Retroviridae	0	216	0
Human immunodeficiency virus	Lentivirus	Retroviridae	0	2354	169
Murine leukemia-related retroviruses	Gammaretrovirus	Retroviridae	1	0	0
Human immunodeficiency virus 1	Lentivirus	Retroviridae	51	52030	3641
Enterobacteria phage HK629	Lambdaliikevirus	Siphoviridae	0	0	4
Propionibacterium phage P1.1		Siphoviridae	0	0	62
Pseudomonas phage D3		Siphoviridae	0	0	90
Enterobacterial phage mEp213	Lambdaliikevirus	Siphoviridae	0	2	0
Enterobacteria phage lambda	Lambdaliikevirus	Siphoviridae	0	2	0
Propionibacterium phage ATCC29399B C		Siphoviridae	0	2	0
Enterobacteria phage mEp460	Lambdaliikevirus	Siphoviridae	0	22	0
Propionibacterium phage PAS50		Siphoviridae	0	2	60
Streptococcus phage 5093		Siphoviridae	0	6	0
Pseudomonas phage F10		Siphoviridae	5	0	0
Pseudomonas phage vB PaeS PMG1		Siphoviridae	90	320	102

Supplemental Table S7. Viral summary table of SURPI alignment to NCBI nt DB for a pediatric diarrhea NGS dataset (B), containing SaV and HPeV-1.

Species	Genus	Family	Index #0 (pool of 16 stool samples)
Environmental Halophage eHP-41			1
Lactococcus phage Q7			1
Southern tomato virus			10
Streptococcus phage 34117			110
Enterobacteria phage CL707			13
Streptococcus phage phi-SsUD.1			14
unidentified phage			14025
Enterobacteria phage HK106			1436
Clostridium phage phiSM101			1499
Streptococcus phage 2167			160
Escherichia phage P13374			168
Acinetobacter bacteriophage AP22			18
Streptococcus phage phi-m46.1			1823
Environmental Halophage eHP-11			2
Geobacillus phage GBSV1			2
Streptococcus phage V22			2
Rhodococcus phage REQ2			23
Escherichia phage ECML-134			237
Streptococcus phage 23782			2413
Klebsiella phage KPP95			276716
Escherichia phage JLK-2012			2836
Salmonella phage ST4			303
Escherichia phage fi125			34
Enterobacteria phage HK542			368
Salmonella phage wksl3			37
Enterobacteria phage VT1-Sakai			4
Lactococcus phage Q42			4
Enterobacteria phage mEp021			46
Cronobacter phage ES2			5
Streptococcus phage 8140			50
Streptococcus phage 040922			55
Streptococcus phage 11865			5675
Lactobacillus phage AQ113			7
Enterobacteria phage HK544			70
Human adenovirus C	Mastadenovirus	Adenoviridae	12
Titi monkey adenovirus ECC-2011		Adenoviridae	290
Cyprinid herpesvirus 1	Cyprinivirus	Alloherpesviridae	7
Torque teno virus 5	Alphatorquevirus	Anelloviridae	1
Torque teno mini virus 1	Betatorquevirus	Anelloviridae	10
Torque teno like mini virus Isfahan		Anelloviridae	11
Torque teno virus		Anelloviridae	1232
Torque teno mini virus 3	Betatorquevirus	Anelloviridae	16
Torque teno mini virus 5	Betatorquevirus	Anelloviridae	1831
Torque teno mini virus 6	Betatorquevirus	Anelloviridae	19
Torque teno midi virus		Anelloviridae	19645
TTV-like mini virus		Anelloviridae	2105
Torque teno mini virus 2	Betatorquevirus	Anelloviridae	25
Torque teno midi virus 2	Gammatorquevirus	Anelloviridae	29
Torque teno virus 28	Alphatorquevirus	Anelloviridae	29
Torque teno mini virus 8	Betatorquevirus	Anelloviridae	35
Small anellovirus		Anelloviridae	423
Torque teno mini virus 9	Betatorquevirus	Anelloviridae	455
Torque teno mini virus 4	Betatorquevirus	Anelloviridae	48
Torque teno mini virus 7	Betatorquevirus	Anelloviridae	63
Torque teno virus 24	Alphatorquevirus	Anelloviridae	9

Torque teno midi virus 1	Gammatorquevirus	Anelloviridae	95
Astrovirus MLB2		Astroviridae	22
Mamastrovirus 1	Mamastrovirus	Astroviridae	32
Epipotia aporema granulovirus	Betabaculovirus	Baculoviridae	1
Maruca vitrata nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	1
Trichoplusia ni single nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	1
Autographa nigrisigna nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	12
Spodoptera litura nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	2
Plutella xylostella multiple nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	25685
Anagrapha falcifera multiple nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	38
Bombyx mori nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	6619
Autographa californica multiple nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	7484
Prunus necrotic ringspot virus	Ilarvirus	Bromoviridae	3
Vesicular exanthema of swine virus	Vesivirus	Caliciviridae	1
Norwalk virus	Norovirus	Caliciviridae	12643
Sapporo virus	Sapovirus	Caliciviridae	378213
Cacao swollen shoot virus	Badnavirus	Caulimoviridae	1
Human gyrovirus type 1	Gyrovirus	Circoviridae	1
Guinea fowl circovirus		Circoviridae	109
Chicken anemia virus	Gyrovirus	Circoviridae	1111
Avian gyrovirus 2	Gyrovirus	Circoviridae	140
Gyrovirus GyV3	Gyrovirus	Circoviridae	1443
Betacoronavirus 1	Betacoronavirus	Coronaviridae	14
Human enteric coronavirus strain 4408		Coronaviridae	2
Bell pepper endornavirus	Endornavirus	Endornaviridae	2
Hepatitis B virus	Orthohepadnavirus	Hepadnaviridae	1
Chimpanzee alpha-1 herpesvirus		Herpesviridae	1
Equid herpesvirus 1	Varicellovirus	Herpesviridae	1
Macacine herpesvirus 3	Cytomegalovirus	Herpesviridae	1
Suid herpesvirus 1	Varicellovirus	Herpesviridae	1
Gallid herpesvirus 2	Mardivirus	Herpesviridae	11
Human herpesvirus 5	Cytomegalovirus	Herpesviridae	11
Human herpesvirus 7	Roseolovirus	Herpesviridae	2
Cercopithecine herpesvirus 5	Cytomegalovirus	Herpesviridae	8
Human herpesvirus 1	Simplexvirus	Herpesviridae	88842
Enterobacteria phage M13	Inovirus	Inoviridae	14883005
Enterobacteria phage fd	Inovirus	Inoviridae	35
Enterobacteria phage f1	Inovirus	Inoviridae	41300
Enterobacteria phage If1	Inovirus	Inoviridae	8
Moumouvirus		Mimiviridae	2
Aeromonas phage 44RR2.8t	T4likevirus	Myoviridae	1
Enterobacteria phage LZ2	T4likevirus	Myoviridae	1
Enterobacteria phage PsP3	P2likeviruses	Myoviridae	1
Enterobacteria phage RB15	T4likevirus	Myoviridae	1
Pseudomonas phage EL	Phikzlikevirus	Myoviridae	1
Serratia phage KSP20		Myoviridae	1
Vibrio phage KVP40	Schizot4likevirus	Myoviridae	1
Erwinia phage phiEt88		Myoviridae	10
Erwinia phage phiEa21-4	Felixounalikevirus	Myoviridae	11
Pseudomonas phage LBL3	Pbunlikevirus	Myoviridae	1161
Enterobacteria phage CC31	T4likevirus	Myoviridae	12784
Enterobacteria phage U5	T4likevirus	Myoviridae	129
Escherichia phage LZ9	T4likevirus	Myoviridae	13
Cyanobacteria phage AS-1		Myoviridae	142
Yersinia phage phiR1-RT		Myoviridae	145
Enterobacteria phage RB32	T4likevirus	Myoviridae	147
Pseudomonas phage JG024		Myoviridae	148
Enterobacteria phage HX01	T4likevirus	Myoviridae	1484
Pseudomonas phage PA26		Myoviridae	155
Enterobacteria phage T4 sensu lato	T4likevirus	Myoviridae	1557
Enterobacteria phage Bp7	T4likevirus	Myoviridae	1647

Enterobacteria phage RB69	T4likevirus	Myoviridae	1660
Pseudomonas phage SN	Pbunlikevirus	Myoviridae	17
Streptococcus phage EJ-1		Myoviridae	18
Enterobacteria phage Tula	T4likevirus	Myoviridae	1809
Enterobacteria phage vB EcoM ACG-C40	T4likevirus	Myoviridae	1840
Aeromonas phage CC2	T4likevirus	Myoviridae	19
Enterobacteria phage RB3	T4likevirus	Myoviridae	2
Pseudomonas phage LMA2	Pbunlikevirus	Myoviridae	2
uncultured Myoviridae		Myoviridae	21
Pseudomonas phage NH-4		Myoviridae	215
Acidithiobacillus phage AcaML1		Myoviridae	228
Aeromonas phage PX29	T4likevirus	Myoviridae	23
Aeromonas phage vB AsaM-56		Myoviridae	23
Acinetobacter phage Acj61	T4likevirus	Myoviridae	234
Klebsiella phage KP27	T4likevirus	Myoviridae	2383
Enterobacteria phage K3	T4likevirus	Myoviridae	24
Enterobacteria phage AR1	T4likevirus	Myoviridae	245
Shigella phage SP18	T4likevirus	Myoviridae	250
Pseudomonas phage PB1	Pbunlikevirus	Myoviridae	263
Clostridium phage phiC2	PhiCD119likevirus	Myoviridae	27
Enterobacteria phage JS10	T4likevirus	Myoviridae	2717
Acinetobacter phage Acj9	T4likevirus	Myoviridae	28
Aeromonas phage phiAS4	T4likevirus	Myoviridae	29
Aeromonas phage 25	T4likevirus	Myoviridae	3
Cronobacter phage CR3		Myoviridae	3
Enterobacteria phage JS98-C3	T4likevirus	Myoviridae	3
Haemophilus phage Aaphi23		Myoviridae	3
Pseudomonas phage 14-1	Pbunlikevirus	Myoviridae	3
Enterobacteria phage P1	Punlikevirus	Myoviridae	3071
Salmonella phage SPN3US		Myoviridae	3308
Cronobacter phage vB CsaM GAP161	T4likevirus	Myoviridae	35
Enterobacteria phage 186	P2likeviruses	Myoviridae	38
Enterobacteria phage SfV		Myoviridae	391
Enterobacteria phage ime09	T4likevirus	Myoviridae	394
Enterobacteria phage IME08	T4likevirus	Myoviridae	400
Enterobacteria phage JS	T4likevirus	Myoviridae	407
Klebsiella phage KP15	T4likevirus	Myoviridae	45
Thermus phage TMA		Myoviridae	45
Acinetobacter phage ZZ1		Myoviridae	456
Enterobacteria phage phiP27		Myoviridae	47
Yersinia phage phiD1		Myoviridae	491
Acinetobacter phage Ac42	T4likevirus	Myoviridae	52
Enterobacteria phage RB51	T4likevirus	Myoviridae	52
Acinetobacter phage 133	T4likevirus	Myoviridae	58
Aeromonas phage Aes508	T4likevirus	Myoviridae	6
Aeromonas phage phiAS5	T4likevirus	Myoviridae	6
Enterobacteria phage phi92		Myoviridae	6
Xanthomonas phage vB XveM DIBBI		Myoviridae	6
Enterobacteria phage vB EcoM-VR7	T4likevirus	Myoviridae	634
Clostridium phage phiMMP04		Myoviridae	67
Yersinia phage L-413C	P2likeviruses	Myoviridae	670
Enterobacteria phage JSE	T4likevirus	Myoviridae	7
Pseudomonas phage KPP12		Myoviridae	719
Enterobacteria phage PP01		Myoviridae	8
Shigella phage Ag3	Viunlikevirus	Myoviridae	8
Shigella phage Shfl2	T4likevirus	Myoviridae	814
Enterobacteria phage RB14	T4likevirus	Myoviridae	83
Cronobacter phage ENT47670		Myoviridae	9
Enterobacteria phage P2	P2likeviruses	Myoviridae	93
Salmonella phage STML-198		Myoviridae	9885
Pseudomonas phage F8	Pbunlikevirus	Myoviridae	99
Escherichia phage wV7	T4likevirus	Myoviridae	990

White spot syndrome virus	Whispovirus	Nimaviridae	1
Influenza A virus	Influenzavirus A	Orthomyxoviridae	8
Human papillomavirus		Papillomaviridae	12
Human respiratory syncytial virus	Pneumovirus	Paramyxoviridae	26604
Respiratory syncytial virus	Pneumovirus	Paramyxoviridae	8840
Mouse adeno-associated virus 1	Dependovirus	Parvoviridae	1
Adeno-associated virus 13	Dependovirus	Parvoviridae	1083
Adeno-associated virus - 2	Dependovirus	Parvoviridae	18
Adeno-associated virus VR-355	Dependovirus	Parvoviridae	19
Non-human primate Adeno-associated virus	Dependovirus	Parvoviridae	2
Adeno-associated virus 12	Dependovirus	Parvoviridae	5
Adeno-associated virus	Dependovirus	Parvoviridae	55033
Human bocavirus	Bocavirus	Parvoviridae	6
Paramecium bursaria Chlorella virus A1	Chlorovirus	Phycodnaviridae	1
Human parechovirus	Parechovirus	Picornaviridae	141266
Enterovirus C	Enterovirus	Picornaviridae	3
Human rhinovirus sp.	Enterovirus	Picornaviridae	4
Rhinovirus A	Enterovirus	Picornaviridae	4
Enterovirus A	Enterovirus	Picornaviridae	45
Enterovirus B	Enterovirus	Picornaviridae	509
Rhinovirus C	Enterovirus	Picornaviridae	5977
Enterobacteria phage W31	T7likevirus	Podoviridae	1
Enterobacter phage EcP1	N4likevirus	Podoviridae	1
Phage PS119		Podoviridae	1
Pseudomonas phage YMC 01 01 P52 PAE BP		Podoviridae	1
Salmonella phage ST64T	P22likevirus	Podoviridae	1
Silicibacter phage DSS3phi2	N4likevirus	Podoviridae	1
Sulfitobacter phage EE36phi1	N4likevirus	Podoviridae	1
Lactococcus phage KSY1		Podoviridae	10
Escherichia phage phiV10	Epsilon15likevirus	Podoviridae	133
Salmonella phage SPN9CC		Podoviridae	144
Escherichia phage vB EcoP G7C	N4likevirus	Podoviridae	16
Yersinia phage phiA1122	T7likevirus	Podoviridae	161496
Salmonella phage epsilon15	Epsilon15likevirus	Podoviridae	1748
Salmonella phage ST160	P22likevirus	Podoviridae	179
Enterobacteria phage 13a	T7likevirus	Podoviridae	1860
Enterobacter phage IME11	N4likevirus	Podoviridae	19
Enterobacteria phage ST104	P22likevirus	Podoviridae	2
Escherichia phage KBNP21	N4likevirus	Podoviridae	2
Pseudomonas phage LUZ7	N4likevirus	Podoviridae	20
Salmonella phage SPN1S		Podoviridae	2612
Erwinia phage PEp14		Podoviridae	28
Pseudomonas phage F116		Podoviridae	405
Salmonella phage Vi06	T7likevirus	Podoviridae	41
Escherichia phage TL-2011c		Podoviridae	42
Salmonella phage HK620	P22likevirus	Podoviridae	438
Salmonella phage vB SemP Emek	P22likevirus	Podoviridae	44
Enterobacteria phage P22	P22likevirus	Podoviridae	49
Erwinia phage vB EamP-S6	N4likevirus	Podoviridae	616
Enterobacteria phage T7	T7likevirus	Podoviridae	66789
Enterobacteria phage CUS-3	P22likevirus	Podoviridae	7
Salmonella phage epsilon34	P22likevirus	Podoviridae	7
Shigella phage Sf6	P22likevirus	Podoviridae	71
Enterobacteria phage T7M	T7likevirus	Podoviridae	732
Salmonella phage g341c	P22likevirus	Podoviridae	8
Escherichia phage TL-2011b	Epsilon15likevirus	Podoviridae	941
Enterobacteria phage IME10		Podoviridae	966
MX polyomavirus	Polyomavirus	Polyomaviridae	320
Wheat streak mosaic virus	Tritimovirus	Potyviridae	13
Onion yellow dwarf virus	Potyvirus	Potyviridae	176
Papaya ringspot virus	Potyvirus	Potyviridae	41
Aquareovirus C	Aquareovirus	Reoviridae	1

Rotavirus A	Rotavirus	Reoviridae	104299
Human rotavirus HRUKM I	Rotavirus	Reoviridae	14
Eubenangee virus	Orbivirus	Reoviridae	2
Rotavirus C	Rotavirus	Reoviridae	39
Moloney murine sarcoma virus	Gammaretrovirus	Retroviridae	1
Murine leukemia-related retroviruses	Gammaretrovirus	Retroviridae	1
Rous sarcoma virus	Alpharetrovirus	Retroviridae	122
Avian leukosis virus	Alpharetrovirus	Retroviridae	150
Murine leukemia virus	Gammaretrovirus	Retroviridae	28
Human immunodeficiency virus 1	Lentivirus	Retroviridae	36
Simian immunodeficiency virus	Lentivirus	Retroviridae	5
Vesicular stomatitis Indiana virus	Vesiculovirus	Rhabdoviridae	3
Recombinant vesicular stomatitis Indiana virus			
rVSV-G GFP	Vesiculovirus	Rhabdoviridae	8
Enterobacteria phage BP-4795	Lambdalikeyvirus	Siphoviridae	1
Lactobacillus phage phiJL-1		Siphoviridae	1
Lactococcus phage bIL310		Siphoviridae	1
Lactococcus phage bIL66M1		Siphoviridae	1
Mycobacterium phage Baka		Siphoviridae	1
Mycobacterium phage Yoshand		Siphoviridae	1
Pseudomonas phage D3		Siphoviridae	1
Riemerella phage RAP44		Siphoviridae	1
Streptococcus phage MM1		Siphoviridae	1
Stx2 converting phage vB EcoP 24B	Lambdalikeyvirus	Siphoviridae	1
Enterobacteria phage T1	Tunalikeyvirus	Siphoviridae	10
Escherichia Stx1 converting bacteriophage	Lambdalikeyvirus	Siphoviridae	10
Salmonella phage SS3e		Siphoviridae	100
Enterobacteria phage BF23	T5likevirus	Siphoviridae	1041
Enterobacteria phage mEp043 c-1	Lambdalikeyvirus	Siphoviridae	1082
Salmonella phage MB78		Siphoviridae	11
Streptococcus phage Sfi21		Siphoviridae	11
Lactococcus phage 949		Siphoviridae	1109
Streptococcus phage SM1		Siphoviridae	1171
Klebsiella phage phiKO2		Siphoviridae	1175
Rhodococcus phage ReqiPoco6		Siphoviridae	126
Streptococcus phage ALQ13.2		Siphoviridae	134
Enterococcus phage phiFL2A		Siphoviridae	14
Lactococcus phage bIL67	C2likevirus	Siphoviridae	14
Propionibacterium phage P100 A		Siphoviridae	14
Escherichia phage bV EcoS AKFV33	T5likevirus	Siphoviridae	141088
Enterobacteria phage HK633	Lambdalikeyvirus	Siphoviridae	145
Clostridium phage phi3626		Siphoviridae	15
Enterobacteria phage HK225	Lambdalikeyvirus	Siphoviridae	1516
Yersinia phage phiR201		Siphoviridae	1545
Staphylococcus phage 187		Siphoviridae	16
Escherichia phage K1-ind (3)		Siphoviridae	1641
Pseudomonas phage phi297		Siphoviridae	166
Enterobacterial phage mEp213	Lambdalikeyvirus	Siphoviridae	1687
Escherichia phage TL-2011a	Lambdalikeyvirus	Siphoviridae	173
Enterobacteria phage mEp460	Lambdalikeyvirus	Siphoviridae	17864
Enterobacteria phage PA-2	Lambdalikeyvirus	Siphoviridae	18
Staphylococcus phage vB SepiS-phiPLA7		Siphoviridae	182
Streptococcus phage Sfi19		Siphoviridae	188
Salmonella phage 5	T5likevirus	Siphoviridae	1939
Escherichia phage EcoS-CEV2	T5likevirus	Siphoviridae	1947
Enterobacteria phage H-19B	Lambdalikeyvirus	Siphoviridae	2
Enterobacteria phage P-EibC		Siphoviridae	2
Escherichia phage ADB-2	Tunalikeyvirus	Siphoviridae	2
Lactococcus phage bIL286		Siphoviridae	2
Lactococcus phage GR6	C2likevirus	Siphoviridae	2
Lactococcus phage TP901-1		Siphoviridae	2
Propionibacterium phage P101A		Siphoviridae	2

Pseudomonas phage JBD88a		Siphoviridae	2
Salmonella phage SETP13		Siphoviridae	2
Enterobacteria phage HK022	Lambdalikevirus	Siphoviridae	20
Propionibacterium phage PA6		Siphoviridae	2050
Streptococcus phage 7201		Siphoviridae	21
Propionibacterium phage P100 1		Siphoviridae	210
Enterobacteria phage mEpX2	Lambdalikevirus	Siphoviridae	214
Enterobacteria phage HK629	Lambdalikevirus	Siphoviridae	2241
Enterobacteria phage T5	T5likevirus	Siphoviridae	22667
Salmonella phage SE2		Siphoviridae	27
Streptococcus phage 2972		Siphoviridae	27
Enterobacterial phage mEp234	Lambdalikevirus	Siphoviridae	2734
Streptococcus phage JX01		Siphoviridae	274
Streptococcus phage Abc2		Siphoviridae	279
Streptococcus phage 858		Siphoviridae	28
Cronobacter phage phiES15		Siphoviridae	29
Lactobacillus phage JCL1032		Siphoviridae	3
Lactococcus phage P001	C2likevirus	Siphoviridae	3
Staphylococcus phage PH15		Siphoviridae	3
Streptococcus phage J1		Siphoviridae	3
Streptococcus phage PH10		Siphoviridae	3
Lactococcus phage c2	C2likevirus	Siphoviridae	30
Salmonella phage vB SenS-Ent1		Siphoviridae	30
Stx2-converting phage 86	Lambdalikevirus	Siphoviridae	30
Pseudomonas phage vB PaeS PMG1		Siphoviridae	301
Pseudomonas phage PAJU2		Siphoviridae	31
Propionibacterium phage PAS50		Siphoviridae	32
Salmonella phage SPN3UB		Siphoviridae	331
Pectobacterium phage My1		Siphoviridae	34
Salmonella phage SETP3		Siphoviridae	35
Lactococcus phage BK5-T		Siphoviridae	354
Streptococcus phage PH15		Siphoviridae	3565
Lactococcus phage r1t		Siphoviridae	357
Enterobacteria phage N15	N15likevirus	Siphoviridae	36
Escherichia phage K1-dep (4)		Siphoviridae	374
Clostridium phage PhiS63		Siphoviridae	38
Enterobacteria phage F20		Siphoviridae	388176
Escherichia phage K1-ind (2)		Siphoviridae	3913
Enterobacteria phage lambda	Lambdalikevirus	Siphoviridae	4
Enterococcus phage phiEf11		Siphoviridae	4
Pseudomonas phage JBD18		Siphoviridae	4
Shigella phage Shf11	Tunalikevirus	Siphoviridae	4
Staphylococcus phage 37		Siphoviridae	41
Streptococcus phage O1205		Siphoviridae	44
Klebsiella phage KP36	Tunalikevirus	Siphoviridae	454297
Enterobacteria phage ES18		Siphoviridae	46
Escherichia phage HK75	Lambdalikevirus	Siphoviridae	477
Enterobacteria phage HK140	Lambdalikevirus	Siphoviridae	4829
Pseudomonas phage JBD67		Siphoviridae	49
Streptococcus phage TP-J34		Siphoviridae	490
Enterobacteria phage TLS	Tunalikevirus	Siphoviridae	5
Pseudomonas phage F10		Siphoviridae	5
Propionibacterium phage P14.4		Siphoviridae	50
Streptococcus phage Sfi11		Siphoviridae	515
Streptococcus phage YMC-2011		Siphoviridae	521
Lactococcus phage P335 sensu lato		Siphoviridae	530
Lactococcus phage phi41		Siphoviridae	539
Enterobacteria phage HK630	Lambdalikevirus	Siphoviridae	54
Pseudomonas phage JBD25		Siphoviridae	54
Streptococcus phage DT1		Siphoviridae	54
Streptococcus phage 5093		Siphoviridae	55330
Enterobacteria phage SPC35	T5likevirus	Siphoviridae	55373

Enterobacteria phage HK446	Lambdaliavirus	Siphoviridae	570
Lactococcus phage Tuc2009		Siphoviridae	59
Enterobacteria phage ST286-24		Siphoviridae	6
Propionibacterium phage PAD20		Siphoviridae	6
Salmonella phage vB SsS Oslo		Siphoviridae	615
Escherichia phage HK639	Lambdaliavirus	Siphoviridae	636
Propionibacterium phage ATCC29399B T		Siphoviridae	64
Escherichia phage K1-ind (1)		Siphoviridae	647
Streptococcus phage phiNJ2		Siphoviridae	66
Enterobacteria phage EPS7		Siphoviridae	6614
Salmonella phage SSU5		Siphoviridae	6628
Propionibacterium phage P1.1		Siphoviridae	73
Enterobacteria phage cdtI	Lambdaliavirus	Siphoviridae	74
Enterobacteria phage HK97	Lambdaliavirus	Siphoviridae	75
Lactococcus phage ul36		Siphoviridae	7500
Lactococcus phage 936 sensu lato		Siphoviridae	7996
Lactococcus phage P087		Siphoviridae	8
Propionibacterium phage P100D		Siphoviridae	8
Streptococcus phage MD2		Siphoviridae	8
Cronobacter phage ENT39118		Siphoviridae	81
Enterobacteria phage mEpX1	Lambdaliavirus	Siphoviridae	810
Enterobacteria phage mEp390	Lambdaliavirus	Siphoviridae	838
Enterobacteria phage mEp237	Lambdaliavirus	Siphoviridae	876
Enterobacteria phage mEp235	Lambdaliavirus	Siphoviridae	8841
Propionibacterium phage P105		Siphoviridae	89
Phage 21	Lambdaliavirus	Siphoviridae	9
Pseudomonas phage B3		Siphoviridae	9
Rhodococcus phage ReqiPepy6		Siphoviridae	902
Erwinia phage phiEaH2		Siphoviridae	921
Propionibacterium phage ATCC29399B C		Siphoviridae	93
Enterobacteria phage D1G		Siphoviridae	94
Propionibacterium phage P104A		Siphoviridae	98
Leuconostoc phage Lmd1		Siphoviridae	99
Escherichia phage K1-dep (1)		Siphoviridae	999
Pepper mild mottle virus	Tobamovirus	Virgaviridae	1
Tomato mosaic virus	Tobamovirus	Virgaviridae	4
HPV 18 Culture			

Supplemental Table S8. Viral summary table of SURPI alignment to NCBI nt DB for a prostate cancer NGS dataset (J), containing HPV-18.

Species	Genus	Family	Index #0 (prostate cancer tissue)
Human herpesvirus 5	Cytomegalovirus	Herpesviridae	1
Murid herpesvirus 1	Muromegalovirus	Herpesviridae	1
Suid herpesvirus 1	Varicellovirus	Herpesviridae	1
Human herpesvirus 1	Simplexvirus	Herpesviridae	2
Alphapapillomavirus 7	Alphapapillomavirus	Papillomaviridae	11955
Human papillomavirus		Papillomaviridae	30
Murine xenotropic virus NZB		Retroviridae	1
Human immunodeficiency virus 1	Lentivirus	Retroviridae	3
Murine leukemia virus	Gammaretrovirus	Retroviridae	31
Murine leukemia-related retroviruses	Gammaretrovirus	Retroviridae	5
Enterobacteria phage phi80	Lambdaliavirus	Siphoviridae	1

Supplemental Table S9. Viral summary table of SURPI alignment to NCBI nt DB for an encephalitis NGS dataset (G), containing HHV-3.

Species	Genus	Family	Index #0 (CSF encephalitis sample)
Lactococcus phage 1358			1
Rice grassy stunt virus	Tenuivirus		1
Agaricus bisporus virus X			14
Environmental Halophage eHP-11			2
Escherichia phage ECML-134			2
Organic Lake virophage			2
Rhodobacter phage RcapNL			2
Salmonella phage SBA-1781			3
Sinorhizobium phage PBC5			6
Environmental Halophage eHP-40			8
Cyprinid herpesvirus 2	Cyprinivirus	Alloherpesviridae	13
Torque teno virus 10	Alphatorquevirus	Anelloviridae	1
Torque teno virus 12	Alphatorquevirus	Anelloviridae	1
Torque teno virus		Anelloviridae	6
Mamastrovirus 1	Mamastrovirus	Astroviridae	2
Murine norovirus	Norovirus	Caliciviridae	1
Tobacco vein clearing virus	Cavemovirus	Caulimoviridae	8
Dengue virus	Flavivirus	Flaviviridae	1
Hepatitis C virus	Hepacivirus	Flaviviridae	1
Tomato leaf curl Taiwan virus	Begomovirus	Geminiviridae	1
Cercopithecine herpesvirus 2	Simplexvirus	Herpesviridae	1
Suid herpesvirus 1	Varicellovirus	Herpesviridae	1
Human herpesvirus 3	Varicellovirus	Herpesviridae	1094
Human herpesvirus 1	Simplexvirus	Herpesviridae	54
Cercopithecine herpesvirus 5	Cytomegalovirus	Herpesviridae	8
Invertebrate iridescent virus 9	Iridovirus	Iridoviridae	129
Invertebrate iridescent virus 16	Iridovirus	Iridoviridae	2
Invertebrate iridescent virus 6	Iridovirus	Iridoviridae	70
Invertebrate iridescent virus 1	Iridovirus	Iridoviridae	8
Enterobacteria phage phiX174	Microvirus	Microviridae	3
Acanthamoeba castellanii mamavirus		Mimiviridae	4
Moumouvirus		Mimiviridae	5
Aeromonas phage vB AsaM-56		Myoviridae	1
Burkholderia phage phiE255	Bcepmlikevirus	Myoviridae	1
Cronobacter phage vB CsaM GAP32		Myoviridae	1
Enterobacteria phage HX01	T4likevirus	Myoviridae	1
Enterobacteria phage T4 sensu lato	T4likevirus	Myoviridae	1
Pseudomonas phage JG024		Myoviridae	1
Pseudomonas phage OBP		Myoviridae	1
Shigella phage Shf12	T4likevirus	Myoviridae	1
Shigella phage SP18	T4likevirus	Myoviridae	1
Burkholderia phage Bups phi1		Myoviridae	14
Pseudomonas phage EL	Phikzlikevirus	Myoviridae	18
Bacillus phage BCD7		Myoviridae	2
Enterobacteria phage vB EcoM ACG-C40	T4likevirus	Myoviridae	2
Pseudomonas phage phiCTX	P2likeviruses	Myoviridae	2
Salmonella phage FelixO1	Felixounalikevirus	Myoviridae	2
Enterobacteria phage CC31	T4likevirus	Myoviridae	3
Escherichia phage wV7	T4likevirus	Myoviridae	3
Halobacterium phage phiH	Phihlikevirus	Myoviridae	3
uncultured Myoviridae		Myoviridae	31
Escherichia phage EC6		Myoviridae	4
Iodobacteriophage phiPLPE		Myoviridae	4
Pseudomonas phage PB1	Pbunlikevirus	Myoviridae	4

Yersinia phage phiD1		Myoviridae	4
Pseudomonas phage KPP12		Myoviridae	45
Synechococcus phage metaG-MbCM1	T4likevirus	Myoviridae	5
Cyanobacteria phage AS-1		Myoviridae	59
Acidithiobacillus phage AcaML1		Myoviridae	830
Adeno-associated virus 13	Dependovirus	Parvoviridae	1
Adeno-associated virus - 2	Dependovirus	Parvoviridae	2
Adeno-associated virus	Dependovirus	Parvoviridae	7
Paramecium bursaria Chlorella virus A1	Chlorovirus	Phycodnaviridae	38
Enterovirus A	Enterovirus	Picornaviridae	1
Enterovirus C	Enterovirus	Picornaviridae	1
Human parechovirus	Parechovirus	Picornaviridae	1
Enterobacteria phage IME10		Podoviridae	1
Salmonella phage 7-11		Podoviridae	1
Salmonella phage phiSG-JL2	T7likevirus	Podoviridae	1
Enterobacter phage IME11	N4likevirus	Podoviridae	10
Brucella phage Pr		Podoviridae	13
Pseudomonas phage phi-2	Phikmvlikevirus	Podoviridae	14
Salmonella phage epsilon15	Epsilon15likevirus	Podoviridae	14
Yersinia phage phiA1122	T7likevirus	Podoviridae	143
Bacillus phage phi29	Phi29likevirus	Podoviridae	22
Salmonella phage epsilon34	P22likevirus	Podoviridae	23
Escherichia phage N4	N4likevirus	Podoviridae	3
Uncultured T7-like podovirus		Podoviridae	3
EBPR podovirus 1		Podoviridae	47
Burkholderia phage Bcep22		Podoviridae	5
Escherichia phage KBNP21	N4likevirus	Podoviridae	5
Bordetella phage BIP-1	Bppunalikevirus	Podoviridae	6
Salinivibrio phage CW02		Podoviridae	6
Enterobacteria phage T7	T7likevirus	Podoviridae	69
Erwinia phage PEp14		Podoviridae	7
EBPR podovirus 3		Podoviridae	73
Salmonella phage g341c	P22likevirus	Podoviridae	8
Enterobacteria phage T7M	T7likevirus	Podoviridae	9
MW polyomavirus	Polyomavirus	Polyomaviridae	9
Avian leukosis virus	Alpharetrovirus	Retroviridae	1
Simian immunodeficiency virus	Lentivirus	Retroviridae	1
Human immunodeficiency virus 1	Lentivirus	Retroviridae	4
Grapevine fanleaf virus	Nepovirus	Secoviridae	4
Burkholderia phage KS9	Lambdalikevirus	Siphoviridae	1
Enterobacterial phage mEp234	Lambdalikevirus	Siphoviridae	1
Enterobacteria phage T5	T5likevirus	Siphoviridae	1
Enterobacteria phage vB EcoS ACG-M12	Tunalikevirus	Siphoviridae	1
Enterococcus phage phiFL3A		Siphoviridae	1
Mycobacterium phage SirDuracell		Siphoviridae	1
Mycobacterium phage Troll4		Siphoviridae	1
EBPR siphovirus 2		Siphoviridae	11
Burkholderia phage KL1		Siphoviridae	12
Enterobacteria phage mEp043 c-1	Lambdalikevirus	Siphoviridae	12
Salmonella phage SSU5		Siphoviridae	13
Propionibacterium phage P100 1		Siphoviridae	15
Pseudomonas phage phi297		Siphoviridae	15
Propionibacterium phage ATCC29399B C		Siphoviridae	18
Enterobacteria phage HK629	Lambdalikevirus	Siphoviridae	195
Mycobacterium phage Charlie		Siphoviridae	2
Pseudomonas phage vB PaeS PMG1		Siphoviridae	2
Salmonella phage SETP12		Siphoviridae	2
Acinetobacter phage Bphi-B1251		Siphoviridae	20
Lactococcus phage TP901-1		Siphoviridae	23
Enterobacteria phage HK630	Lambdalikevirus	Siphoviridae	25
Enterobacteria phage HK140	Lambdalikevirus	Siphoviridae	3
Enterobacteria phage HK578		Siphoviridae	3

Escherichia phage K1-ind (2)		Siphoviridae	3
Mycobacterium phage Spartacus		Siphoviridae	3
Pseudomonas phage vB Pae-Kakheti25		Siphoviridae	3
Mycobacterium phage Pacc40		Siphoviridae	34
Lactococcus phage P335 sensu lato		Siphoviridae	4
Mycobacterium phage Avrafan		Siphoviridae	4
Propionibacterium phage PAS50		Siphoviridae	4
Vibrio phage pVp-1	T5likevirus	Siphoviridae	4
Phage phiJL001	Yualikevirus	Siphoviridae	44
Lactococcus phage 936 sensu lato		Siphoviridae	5
Mycobacterium phage Firecracker		Siphoviridae	5
Propionibacterium phage P100 A		Siphoviridae	5
Propionibacterium phage P9.1		Siphoviridae	5
Escherichia phage K1-ind (1)		Siphoviridae	6
Lactococcus phage BK5-T		Siphoviridae	7
Propionibacterium phage P1.1		Siphoviridae	8
Mycobacterium phage Brujita		Siphoviridae	9
Pseudomonas phage 73		Siphoviridae	9

Supplemental Table S10. Viral summary table of SURPI alignment to NCBI nt DB for a TTVS hepatitis NGS dataset (E), containing HCV-1b.

Species	Genus	Family	Index #CACGAT (HCV-positive serum sample)
Enterobacteria phage HK544			10
Environmental Halophage eHP-11			4
Enterobacteria phage HK542			6
Heliothis zea virus 1			6
Cyprinid herpesvirus 1	Cyprinivirus	Alloherpesviridae	1
Pepino mosaic virus	Potexvirus	Alphaflexiviridae	131
Small anellovirus		Anelloviridae	1
Torque teno midi virus		Anelloviridae	1
Torque teno mini virus 8	Betatorquevirus	Anelloviridae	1
Torque teno virus 15	Alphatorquevirus	Anelloviridae	1
Torque teno virus 21	Alphatorquevirus	Anelloviridae	1
Torque teno virus 3	Alphatorquevirus	Anelloviridae	1
uncultured Anellovirus		Anelloviridae	1
SEN virus		Anelloviridae	10
Torque teno virus 14	Alphatorquevirus	Anelloviridae	11
Torque teno virus		Anelloviridae	2115
TTV-like mini virus		Anelloviridae	4
Torque teno virus 10	Alphatorquevirus	Anelloviridae	6
Torque teno virus 19	Alphatorquevirus	Anelloviridae	8
Plutella xylostella multiple nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	2
Bombyx mori nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	6
GB virus C	Pegivirus	Flaviviridae	1558199
Hepatitis C virus	Hepacivirus	Flaviviridae	186038
Hepatitis B virus	Orthohepadnavirus	Hepadnaviridae	1
Equid herpesvirus 1	Varicellovirus	Herpesviridae	1
Meleagrid herpesvirus 1	Mardivirus	Herpesviridae	1
Macacine herpesvirus 3	Cytomegalovirus	Herpesviridae	2
Cercopithecine herpesvirus 5	Cytomegalovirus	Herpesviridae	73
Enterobacteria phage M13	Inovirus	Inoviridae	157
Enterobacteria phage G4	Microvirus	Microviridae	1
Enterobacteria phage WA13 sensu lato	Microvirus	Microviridae	1
Enterobacteria phage phiX174	Microvirus	Microviridae	31352
Enterobacteria phage phiX174 sensu lato	Microvirus	Microviridae	921
Megavirus courdo7		Mimiviridae	1
Acanthamoeba polyphaga mimivirus	Mimivirus	Mimiviridae	2
Acanthamoeba castellanii mamavirus		Mimiviridae	3
Megavirus chilensis		Mimiviridae	3
Moumouvirus		Mimiviridae	4
Cyanobacteria phage AS-1		Myoviridae	11
Pectobacterium phage phiTE		Myoviridae	115
Acinetobacter phage 133	T4likevirus	Myoviridae	2
Enterobacteria phage JS10	T4likevirus	Myoviridae	2
Cronobacter phage CR3		Myoviridae	38
Acidithiobacillus phage AcaML1		Myoviridae	42
Listeria phage A511	Twortlikevirus	Myoviridae	7
Enterobacteria phage T7M	T7likevirus	Podoviridae	1
Vibrio phage phiVC8		Podoviridae	1
Enterobacteria phage T7	T7likevirus	Podoviridae	1032
Salmonella phage Vi06	T7likevirus	Podoviridae	14
Enterobacteria phage 13a	T7likevirus	Podoviridae	3
Yersinia phage phiA1122	T7likevirus	Podoviridae	700
Simian virus 40	Polyomavirus	Polyomaviridae	1
Turnip mosaic virus	Potyvirus	Potyviridae	1

Human immunodeficiency virus 1	Lentivirus	Retroviridae	1
Enterobacterial phage mEp234	Lambdaliikevirus	Siphoviridae	1
Enterobacteria phage lambda	Lambdaliikevirus	Siphoviridae	1
Propionibacterium phage ATCC29399B C		Siphoviridae	1
Pseudomonas phage B3		Siphoviridae	1
Enterobacterial phage mEp213	Lambdaliikevirus	Siphoviridae	106
Enterobacteria phage HK629	Lambdaliikevirus	Siphoviridae	110
Enterobacteria phage mEp460	Lambdaliikevirus	Siphoviridae	16
Enterobacteria phage mEp043 c-1	Lambdaliikevirus	Siphoviridae	173
Enterobacteria phage HK630	Lambdaliikevirus	Siphoviridae	3
Tomato mosaic virus	Tobamovirus	Virgaviridae	407

Supplemental Table S11. Viral summary table of SURPI alignment to NCBI nt DB for a simian pneumonia NGS dataset (K), containing TMAV.

Species	Genus	Family	Index #0 (lung swab and tissue)
Cynomolgus adenovirus 1	Mastadenovirus	Adenoviridae	1
Gorilla gorilla beringei adenovirus 6	Mastadenovirus	Adenoviridae	1
Macaca fascicularis adenovirus 1	Mastadenovirus	Adenoviridae	1
Macaca mulatta adenovirus 1	Mastadenovirus	Adenoviridae	1
Mouse adenovirus 2 strain K87	Mastadenovirus	Adenoviridae	1
Ovine adenovirus B	Mastadenovirus	Adenoviridae	1
Ovine adenovirus C	Mastadenovirus	Adenoviridae	1
Pan troglodytes schweinfurthii adenovirus 6	Mastadenovirus	Adenoviridae	1
Simian adenovirus 31	Mastadenovirus	Adenoviridae	1
Porcine adenovirus A	Mastadenovirus	Adenoviridae	10
Simian adenovirus 45	Mastadenovirus	Adenoviridae	10
Pan troglodytes verus adenovirus 5	Mastadenovirus	Adenoviridae	11
Simian adenovirus 18	Mastadenovirus	Adenoviridae	129
Human adenovirus F	Mastadenovirus	Adenoviridae	13
Human adenovirus E	Mastadenovirus	Adenoviridae	148
Murine adenovirus B	Mastadenovirus	Adenoviridae	15
Pan troglodytes schweinfurthii adenovirus 11	Mastadenovirus	Adenoviridae	16
Simian adenovirus B	Mastadenovirus	Adenoviridae	160
Human adenovirus C	Mastadenovirus	Adenoviridae	19
Gorilla gorilla gorilla adenovirus 1	Mastadenovirus	Adenoviridae	2
Human adenovirus A	Mastadenovirus	Adenoviridae	2
Ovine adenovirus A	Mastadenovirus	Adenoviridae	2
Pan troglodytes schweinfurthii adenovirus 9	Mastadenovirus	Adenoviridae	2
Porcine adenovirus C	Mastadenovirus	Adenoviridae	2
Simian adenovirus 33	Mastadenovirus	Adenoviridae	2
Simian adenovirus 41	Mastadenovirus	Adenoviridae	2
Simian adenovirus 43	Mastadenovirus	Adenoviridae	2
Simian adenovirus 46	Mastadenovirus	Adenoviridae	2
unidentified adenovirus	Mastadenovirus	Adenoviridae	2
Human adenovirus D	Mastadenovirus	Adenoviridae	254
Bat adenovirus A	Mastadenovirus	Adenoviridae	3
Simian adenovirus 20	Mastadenovirus	Adenoviridae	3
Simian adenovirus 42	Mastadenovirus	Adenoviridae	3
Human adenovirus G	Mastadenovirus	Adenoviridae	32
Piliocolobus badius adenovirus 2	Mastadenovirus	Adenoviridae	37
Simian adenovirus A	Mastadenovirus	Adenoviridae	37
Simian adenovirus 40	Mastadenovirus	Adenoviridae	39
Pan troglodytes schweinfurthii adenovirus 14	Mastadenovirus	Adenoviridae	4
Piliocolobus badius adenovirus 3	Mastadenovirus	Adenoviridae	4
Colobus guereza adenovirus 1	Mastadenovirus	Adenoviridae	5
Human adenovirus B	Mastadenovirus	Adenoviridae	64
Simian adenovirus 7	Mastadenovirus	Adenoviridae	7
Equine adenovirus A	Mastadenovirus	Adenoviridae	8
Saguinus adenovirus 1	Mastadenovirus	Adenoviridae	8
Cyprinid herpesvirus 1	Cyprinivirus	Alloherpesviridae	1
Pepino mosaic virus	Potexvirus	Alphaflexiviridae	50
Torque teno virus		Anelloviridae	2
Trichoplusia ni ascovirus 2c	Ascovirus	Ascoviridae	1
California sea lion astrovirus 5		Astroviridae	1
Mamastrovirus 1	Mamastrovirus	Astroviridae	6
Plutella xylostella multiple nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	4
Autographa californica multiple nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	5
Bombyx mori nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	5
Aguacate virus	Phlebovirus	Bunyaviridae	1

Sapporo virus	Sapovirus	Caliciviridae	11
Human herpesvirus 7	Roseolovirus	Herpesviridae	1
Human herpesvirus 8	Rhadinovirus	Herpesviridae	1
Murid herpesvirus 4	Rhadinovirus	Herpesviridae	1
Human herpesvirus 1	Simplexvirus	Herpesviridae	20
Cercopithecine herpesvirus 5	Cytomegalovirus	Herpesviridae	4
Trichoderma hypovirus	Hypovirus	Hypoviridae	2
Enterobacteria phage f1	Inovirus	Inoviridae	2
Enterobacteria phage M13	Inovirus	Inoviridae	433
Staphylococcus phage Twort	Twortlikevirus	Myoviridae	1
uncultured Myoviridae		Myoviridae	1
Influenza A virus	Influenzavirus A	Orthomyxoviridae	2
Alphapapillomavirus 9	Alphapapillomavirus	Papillomaviridae	1
Hamster parvovirus	Parvovirus	Parvoviridae	1
Enterovirus E	Enterovirus	Picornaviridae	1
Simian enterovirus SV46	Enterovirus	Picornaviridae	1
Enterovirus A	Enterovirus	Picornaviridae	32
WUHARV Enterovirus 3	Enterovirus	Picornaviridae	37
Enterovirus B	Enterovirus	Picornaviridae	4
WUHARV Enterovirus 2	Enterovirus	Picornaviridae	4
WUHARV Enterovirus 1	Enterovirus	Picornaviridae	6
Enterovirus J	Enterovirus	Picornaviridae	7
Escherichia phage vB EcoP G7C	N4likevirus	Podoviridae	1
Yersinia phage phiA1122	T7likevirus	Podoviridae	1
Pseudomonas phage LUZ24	Luz24likevirus	Podoviridae	6
Cotesia sesamiae bracovirus	Bracovirus	Polydnaviridae	1
Rotavirus A	Rotavirus	Reoviridae	186
Feline leukemia virus	Gammaretrovirus	Retroviridae	1
Jaagsiekte sheep retrovirus	Betaretrovirus	Retroviridae	1
Human immunodeficiency virus 1	Lentivirus	Retroviridae	2
Murine leukemia virus	Gammaretrovirus	Retroviridae	3
Enterobacteria phage mEp043 c-1	Lambdalikeyvirus	Siphoviridae	1
Escherichia phage K1-dep (4)		Siphoviridae	1
Lactobacillus phage phiadh		Siphoviridae	1
Lactococcus phage phi41		Siphoviridae	1
Lactococcus phage phiUS3		Siphoviridae	1
Lactococcus phage TP901-1		Siphoviridae	1
Propionibacterium phage P100 1		Siphoviridae	1
Propionibacterium phage P100 A		Siphoviridae	1
Propionibacterium phage P105		Siphoviridae	1
Propionibacterium phage PAD20		Siphoviridae	1
Salmonella phage SE2		Siphoviridae	1
Streptococcus phage TP-J34		Siphoviridae	1
Propionibacterium phage PAS50		Siphoviridae	2
Escherichia phage K1-ind (2)		Siphoviridae	3
Propionibacterium phage PA6		Siphoviridae	3
Salmonella phage Vi II-E1		Siphoviridae	32
Enterobacteria phage HK629	Lambdalikeyvirus	Siphoviridae	38
Propionibacterium phage ATCC29399B T		Siphoviridae	4
Lactococcus phage 936 sensu lato		Siphoviridae	51
Propionibacterium phage ATCC29399B C		Siphoviridae	6

Supplemental Table S12. Viral summary table of SURPI alignment to NCBI nt DB for a hemorrhagic fever NGS dataset (D), containing BASV.

Species	Genus	Family	hemorrhagic fever serum sample
Human herpesvirus 1	Simplexvirus	Herpesviridae	2
Enterobacteria phage f1	Inovirus	Inoviridae	1
Enterobacteria phage M13	Inovirus	Inoviridae	1883
Enterobacteria phage phiX174	Microvirus	Microviridae	1
Burkholderia phage KS5	P2likeviruses	Myoviridae	1
Burkholderia phage phiE202	P2likeviruses	Myoviridae	1
Ralstonia phage RSA1	P2likeviruses	Myoviridae	1
Burkholderia phage KS14	P2likeviruses	Myoviridae	3
Burkholderia phage phiE255	Bcepmlikevirus	Myoviridae	3
Enterobacteria phage T4 sensu lato	T4likevirus	Myoviridae	3
Burkholderia phage KL3	P2likeviruses	Myoviridae	31
Acidithiobacillus phage AcaML1		Myoviridae	6
Burkholderia phage phi52237	P2likeviruses	Myoviridae	7
Pseudomonas phage KPP12		Myoviridae	842
Acanthocystis turfacea Chlorella virus 1	Chlorovirus	Phycodnaviridae	1
Paramecium bursaria Chlorella virus AR158	Chlorovirus	Phycodnaviridae	1
Enterobacteria phage T7	T7likevirus	Podoviridae	1
Pseudomonas phage F116		Podoviridae	1
Burkholderia phage BcepC6B	Bppunlikevirus	Podoviridae	2
Yersinia phage phiA1122	T7likevirus	Podoviridae	8
Rotavirus A	Rotavirus	Reoviridae	33214
Murine leukemia virus	Gammaretrovirus	Retroviridae	12
Burkholderia phage phi644-2	Lambdaliikevirus	Siphoviridae	1
Caulobacter phage CcrRogue		Siphoviridae	1
Propionibacterium phage P1.1		Siphoviridae	1
Mycobacterium phage Jebeks		Siphoviridae	18
Mycobacterium phage BigNuz		Siphoviridae	3
Propionibacterium phage P105		Siphoviridae	4
Propionibacterium phage P14.4		Siphoviridae	5
Burkholderia phage phiE125	Lambdaliikevirus	Siphoviridae	53
Tomato mosaic virus	Tobamovirus	Virgaviridae	4

Supplemental Table S13. Viral summary table of SURPI alignment to NCBI nt DB for a hemorrhagic fever NGS dataset (H), containing BASV.

Species	Genus	Family	hemorrhagic fever serum sample
Enterobacteria phage HK542			11
Enterobacteria phage HK544			2
Environmental Halophage eHP-11			2
Anguillid herpesvirus 1	Cyprinivirus	Alloherpesviridae	1
Cyprinid herpesvirus 2	Cyprinivirus	Alloherpesviridae	1
Cyprinid herpesvirus 1	Cyprinivirus	Alloherpesviridae	11
Torque teno mini virus 9	Betatorquevirus	Anelloviridae	1
TTV-like mini virus		Anelloviridae	1
Torque teno midi virus		Anelloviridae	11
Torque teno mini virus 2	Betatorquevirus	Anelloviridae	2
Zaire ebolavirus	Ebolavirus	Filoviridae	27
Gallid herpesvirus 3	Mardivirus	Herpesviridae	1
Human herpesvirus 6B	Roseolovirus	Herpesviridae	1
Macacine herpesvirus 3	Cytomegalovirus	Herpesviridae	1
Gallid herpesvirus 2	Mardivirus	Herpesviridae	9
Enterobacteria phage M13	Inovirus	Inoviridae	33
Acidithiobacillus phage AcaML1		Myoviridae	2
Emiliana huxleyi virus 99B1	Coccolithovirus	Phycodnaviridae	1
Enterobacteria phage 13a	T7likevirus	Podoviridae	1
Enterobacteria phage T7M	T7likevirus	Podoviridae	2
Salmonella phage Vi06	T7likevirus	Podoviridae	30
Yersinia phage phiA1122	T7likevirus	Podoviridae	348
Enterobacteria phage T7	T7likevirus	Podoviridae	437
Simian virus 40	Polyomavirus	Polyomaviridae	1
Human endogenous retrovirus K		Retroviridae	1
Propionibacterium phage P100 1		Siphoviridae	1
Propionibacterium phage P100 A		Siphoviridae	1
Propionibacterium phage P101A		Siphoviridae	1
Propionibacterium phage P104A		Siphoviridae	1
Pseudomonas phage PAJU2		Siphoviridae	1
Propionibacterium phage P100D		Siphoviridae	10
Klebsiella phage KP36	Tunalikeyvirus	Siphoviridae	14
Salmonella phage SPN3UB		Siphoviridae	15
Enterobacteria phage HK629	Lambdalikeyvirus	Siphoviridae	17
Propionibacterium phage PA6		Siphoviridae	19
Enterobacteria phage HK630	Lambdalikeyvirus	Siphoviridae	2
Propionibacterium phage P14.4		Siphoviridae	2
Pseudomonas phage D3		Siphoviridae	2
Propionibacterium phage ATCC29399B T		Siphoviridae	4
Propionibacterium phage P1.1		Siphoviridae	4
Propionibacterium phage PAS50		Siphoviridae	5
Enterobacteria phage mEp460	Lambdalikeyvirus	Siphoviridae	6
Propionibacterium phage ATCC29399B C		Siphoviridae	66
Enterobacteria phage mEp043 c-1	Lambdalikeyvirus	Siphoviridae	7
Staphylococcus phage StB27		Siphoviridae	7
Propionibacterium phage PAD20		Siphoviridae	9
Pseudomonas phage vB PaeS PMG1		Siphoviridae	9

Supplemental Tables S14-S21. Viral species summary tables of SURPI (RAPSearch) translated nucleotide alignment to the viral protein DB. Letters or numbers following the hash mark denote multiplexed index barcodes. Abbreviations: Contigs, contiguous sequences; HIV-1, human immunodeficiency type 1; SaV, sapovirus; HPeV-1, human parechovirus type 1; BASV, Bas-Congo virus; HCV-1b, hepatitis C virus, type 1b; HHV-3, human herpesvirus type 3; HPV-18, human papillomavirus virus type 18; TMAv, titi monkey adenovirus.

Supplemental Table S14. Viral output of translated nucleotide alignment to viral protein DB for a spiked HIV NGS dataset (L), containing HIV-1.

Species	Genus	Family	Contigs #1 (10 ² HIV)	Reads #1	Contigs #2 (10 ⁴ HIV)	Reads #2	Contigs #3 (10 ³ HIV)	Reads #3
Circovirus-like genome RW-E			0	0	0	9	0	0
Heliothis zea virus 1			0	3	0	0	0	0
Cyprinid herpesvirus 3	Cyprinivirus	Alloherpesviridae	0	0	0	1	0	0
Ranid herpesvirus 2	Batrachovirus	Alloherpesviridae	0	1	0	0	0	0
Torque teno virus 27	Alphatorquevirus	Anelloviridae	0	0	0	0	0	1
Torque teno virus 29	Alphatorquevirus	Anelloviridae	0	0	0	1	0	0
Torque teno mini virus 6	Betatorquevirus	Anelloviridae	0	0	0	20	0	0
Torque teno virus 10	Alphatorquevirus	Anelloviridae	0	1	0	0	0	0
Torque teno mini virus 4	Betatorquevirus	Anelloviridae	0	10	0	0	0	0
Torque teno virus 4	Alphatorquevirus	Anelloviridae	0	10	0	0	0	0
Torque teno midi virus 2	Gammatorquevirus	Anelloviridae	0	10	0	9	0	0
Torque teno midi virus		Anelloviridae	0	10	0	9	0	2
Torque teno virus 20	Alphatorquevirus	Anelloviridae	0	1	0	14	0	0
Torque teno midi virus 1	Gammatorquevirus	Anelloviridae	0	2	0	0	0	5
Small anellovirus		Anelloviridae	0	20	0	14	0	0
Torque teno virus 23	Alphatorquevirus	Anelloviridae	0	3	0	0	0	0
SEN virus		Anelloviridae	0	4	0	0	0	0
Torque teno virus 13	Alphatorquevirus	Anelloviridae	0	4	0	0	0	1
Torque teno virus 12	Alphatorquevirus	Anelloviridae	0	6	0	0	0	0
Torque teno virus 2	Alphatorquevirus	Anelloviridae	0	9	0	0	0	0
Torque teno virus 11	Alphatorquevirus	Anelloviridae	1	10	0	0	0	0
TTV-like mini virus		Anelloviridae	1	16	0	6	0	4
Torque teno virus		Anelloviridae	5	180	0	14	0	5
Spodoptera frugiperda ascovirus 1a	Ascovirus	Ascoviridae	0	0	0	1	0	0
Mamastrovirus 1	Mamastrovirus	Astroviridae	0	0	0	0	0	3
Autographa californica multiple nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	0	0	0	1	0	0
Norwalk virus	Norovirus	Caliciviridae	0	0	0	0	0	5
Banana streak UL virus	Badnavirus	Caulimoviridae	0	0	0	1	0	0
Mirabilis mosaic virus	Caulimovirus	Caulimoviridae	0	0	0	1	0	0
Stilbocarpa mosaic bacilliform virus	Badnavirus	Caulimoviridae	0	0	0	1	0	0
Cacao swollen shoot virus	Badnavirus	Caulimoviridae	0	0	0	10	0	0

Bougainvillea spectabilis chlorotic vein-banding virus	Badnavirus	Caulimoviridae	0	0	0	14	0	0
Lamium leaf distortion virus	Caulimovirus	Caulimoviridae	0	0	0	2	0	0
Pineapple bacilliform virus	Badnavirus	Caulimoviridae	0	0	0	2	0	0
Dioscorea bacilliform virus	Badnavirus	Caulimoviridae	0	0	0	4	0	0
Ambrosia asymptomatic virus 2 UKM-2007	Badnavirus	Caulimoviridae	0	0	0	41	0	4
Beak and feather disease virus	Circovirus	Circoviridae	0	0	0	1	0	0
Pegivirus A	Pegivirus	Flaviviridae	0	0	0	0	0	1
Bovine viral diarrhea virus 1	Pestivirus	Flaviviridae	0	0	0	1	0	0
Hepatitis C virus	Hepacivirus	Flaviviridae	0	0	0	5	0	0
GB virus C	Pegivirus	Flaviviridae	9	109	8	298	1	237
Felid herpesvirus 1	Varicellovirus	Herpesviridae	0	0	0	1	0	0
Gallid herpesvirus 2	Mardivirus	Herpesviridae	0	1	0	0	0	0
Human herpesvirus 8	Rhadinovirus	Herpesviridae	0	4	0	2	0	0
Glossina hytovirus	Glossinavirus	Hytrosaviridae	0	0	0	2	0	0
Musca hytovirus	Muscavirus	Hytrosaviridae	0	0	0	2	0	1
Invertebrate iridescent virus 16	Iridovirus	Iridoviridae	0	0	0	0	0	2
Invertebrate iridescent virus 3	Chloriridovirus	Iridoviridae	0	0	0	0	0	2
Invertebrate iridescent virus 6	Iridovirus	Iridoviridae	0	0	0	0	0	2
Frog virus 3	Ranavirus	Iridoviridae	0	0	0	1	0	0
Moumouvirus		Mimiviridae	0	0	0	0	0	2
Terra1 virus		Mimiviridae	0	14	0	0	0	0
Acanthamoeba polyphaga mimivirus	Mimivirus	Mimiviridae	0	8	0	9	0	15
Feldmannia irregularis virus a	Phaeovirus	Phycodnaviridae	0	0	0	0	0	2
Paramecium bursaria	Chlorovirus	Phycodnaviridae	0	0	0	1	0	0
Chlorella virus 1								
Paramecium bursaria								
Chlorella virus	Chlorovirus	Phycodnaviridae	0	0	0	1	0	0
Marburg 1								
Micromonas sp. RCC1109 virus MpV1	Prasinovirus	Phycodnaviridae	0	0	0	6	0	0
Emiliania huxleyi virus 99B1	Coccolithovirus	Phycodnaviridae	0	0	0	8	0	0
Ostreococcus virus OsV5		Phycodnaviridae	0	1	0	0	0	0
Emiliania huxleyi virus 86	Coccolithovirus	Phycodnaviridae	0	2	0	1	0	0
Ostreococcus lucimarinus virus OIV1	Prasinovirus	Phycodnaviridae	0	2	0	3	0	0
Bathycoccus sp. RCC1105 virus BpV	Prasinovirus	Phycodnaviridae	0	3	0	2	0	8
Feldmannia species virus	Phaeovirus	Phycodnaviridae	0	4	0	0	0	0

Paramecium bursaria Chlorella virus AR158	Chlorovirus	Phycodnaviridae	0	6	0	1	0	0
Paramecium bursaria Chlorella virus NY2A	Chlorovirus	Phycodnaviridae	0	6	0	4	0	1
Paramecium bursaria Chlorella virus A1	Chlorovirus	Phycodnaviridae	0	6	0	8	0	8
Acanthocystis turfacea Chlorella virus 1	Chlorovirus	Phycodnaviridae	1	30	0	4	1	9
Cotesia congregata bracovirus	Bracovirus	Polydnaviridae	0	1	0	0	0	1
Ugandan cassava brown streak virus	Ipomovirus	Potyviridae	0	0	0	1	0	0
Mouse mammary tumor virus	Betaretrovirus	Retroviridae	0	0	0	0	0	1
Human immunodeficiency virus	Lentivirus	Retroviridae	0	0	0	1	0	0
Simian immunodeficiency virus	Lentivirus	Retroviridae	0	0	0	1	0	0
Caprine arthritis encephalitis virus	Lentivirus	Retroviridae	0	0	0	3	0	0
Human immunodeficiency virus 1	Lentivirus	Retroviridae	0	6	2	201	2	20

Supplemental Table S15. Viral output of translated nucleotide alignment to viral protein DB for a pediatric diarrhea NGS dataset (B), containing SaV and HPeV-1.

Species	Genus	Family	Contigs #0	Reads #0
Diaporthe ambigua RNA virus 1			0	1
Heliothis zea virus 1			0	106
Sclerotinia sclerotiorum hypovirulence associated DNA virus 1			0	181
Oryctes rhinoceros virus			0	232
Gryllus bimaculatus nudivirus			0	4
Mavirus			0	4
Cryptosporidium dsRNA virus			0	5
White ash mosaic virus			1	0
Pseudomonas phage PPA5			1	12
Abalone shriveling syndrome-associated virus			3	12
Human adenovirus A	Mastadenovirus	Adenoviridae	0	1
Simian adenovirus A	Mastadenovirus	Adenoviridae	0	2
Turkey adenovirus A	Siadenovirus	Adenoviridae	0	2
Human adenovirus C	Mastadenovirus	Adenoviridae	1	1
Ranid herpesvirus 2	Batrachovirus	Alloherpesviridae	0	50
Cyprinid herpesvirus 3	Cyprinivirus	Alloherpesviridae	0	6
Anguillid herpesvirus 1	Cyprinivirus	Alloherpesviridae	0	7
Ictalurid herpesvirus 1	Ictalurivirus	Alloherpesviridae	1	0
Torque teno douroucouli virus	Zetatorquevirus	Anelloviridae	0	1
Torque teno sus virus 1b	lotatorquevirus	Anelloviridae	0	1
Torque teno virus 1	Alphatorquevirus	Anelloviridae	0	1
Torque teno virus 8	Alphatorquevirus	Anelloviridae	0	1
Torque teno sus virus 1a	lotatorquevirus	Anelloviridae	0	11
Torque teno virus 15	Alphatorquevirus	Anelloviridae	0	11
Torque teno virus 24	Alphatorquevirus	Anelloviridae	0	11
SEN virus		Anelloviridae	0	12
Torque teno mini virus 1	Betatorquevirus	Anelloviridae	0	15
Torque teno virus 21	Alphatorquevirus	Anelloviridae	0	2
Torque teno virus 28	Alphatorquevirus	Anelloviridae	0	21
Torque teno midi virus 1	Gammatorquevirus	Anelloviridae	0	28
Torque teno virus 7	Alphatorquevirus	Anelloviridae	0	3
Torque teno mini virus 7	Betatorquevirus	Anelloviridae	0	354
Torque teno virus 27	Alphatorquevirus	Anelloviridae	0	4
Torque teno virus 29	Alphatorquevirus	Anelloviridae	0	46
Torque teno virus 5	Alphatorquevirus	Anelloviridae	0	5
Torque teno virus 25	Alphatorquevirus	Anelloviridae	0	8
Torque teno mini virus 5	Betatorquevirus	Anelloviridae	0	9
Torque teno virus 23	Alphatorquevirus	Anelloviridae	1	12
Torque teno mini virus 9	Betatorquevirus	Anelloviridae	1	165
Torque teno midi virus 2	Gammatorquevirus	Anelloviridae	1	18
Torque teno virus 26	Alphatorquevirus	Anelloviridae	1	7
Torque teno midi virus		Anelloviridae	20	6715
Torque teno mini virus 6	Betatorquevirus	Anelloviridae	2	254
TTV-like mini virus		Anelloviridae	24	9761
Torque teno virus 13	Alphatorquevirus	Anelloviridae	2	54
Torque teno mini virus 3	Betatorquevirus	Anelloviridae	3	347
Torque teno virus		Anelloviridae	3	377
Torque teno mini virus 2	Betatorquevirus	Anelloviridae	3	509
Torque teno mini virus 4	Betatorquevirus	Anelloviridae	4	692
Small anellovirus		Anelloviridae	5	273
Torque teno mini virus 8	Betatorquevirus	Anelloviridae	6	315
Tacaribe virus	Arenavirus	Arenaviridae	0	1
Equine arteritis virus	Arterivirus	Arteriviridae	0	1

Trichoplusia ni ascovirus 2c	Ascovirus	Ascoviridae	0	18
Heliothis virescens ascovirus 3a	Ascovirus	Ascoviridae	0	19
Diadromus pulchellus ascovirus 4a	Ascovirus	Ascoviridae	3	164
African swine fever virus	Asfivirus	Asfarviridae	0	166
Mamastrovirus 1	Mamastrovirus	Astroviridae	1	4
Choristoneura fumiferana DEF multiple nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	0	1
Cryptophlebia leucotreta granulovirus	Betabaculovirus	Baculoviridae	0	1
Epiphyas postvittana nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	0	1
Helicoverpa armigera granulovirus	Betabaculovirus	Baculoviridae	0	1
Helicoverpa armigera multiple nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	0	1
Helicoverpa armigera nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	0	1
Leucania separata nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	0	1
Plutella xylostella granulovirus	Betabaculovirus	Baculoviridae	0	1
Xestia c-nigrum granulovirus	Betabaculovirus	Baculoviridae	0	1
Epinotia aporema granulovirus	Betabaculovirus	Baculoviridae	0	2
Chrysodeixis chalcites nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	0	3
Mamestra configurata nucleopolyhedrovirus B	Alphabaculovirus	Baculoviridae	0	3
Orgyia pseudotsugata multiple nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	0	3
Hyphantria cunea nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	0	4
Mamestra configurata nucleopolyhedrovirus A	Alphabaculovirus	Baculoviridae	0	4
Agrotis segetum granulovirus	Betabaculovirus	Baculoviridae	0	5
Agrotis segetum nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	0	5
Culex nigripalpus nucleopolyhedrovirus	Deltabaculovirus	Baculoviridae	0	7
Maruca vitrata nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	0	7
Orgyia leucostigma NPV	Alphabaculovirus	Baculoviridae	0	8
Spodoptera litura nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	0	8
Antheraea pernyi nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	1	0
Antheraea polyphemus nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	1	0
Ectropis obliqua nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	1	0
Pieris rapae granulovirus	Betabaculovirus	Baculoviridae	1	0
Anticarsia gemmatilis nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	1	1
Choristoneura fumiferana multiple nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	1	9
Lymantria xyliina MNPV	Alphabaculovirus	Baculoviridae	2	165
Bombyx mori nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	2	22
Autographa californica multiple nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	23	220
Plutella xylostella multiple nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	4	16
Trichoplusia ni single nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	5	877
Tulare apple mosaic virus	Illavirus	Bromoviridae	0	1
Tobacco streak virus	Illavirus	Bromoviridae	0	3
Bunyamwera virus	Orthobunyavirus	Bunyaviridae	0	1
Groundnut bud necrosis virus	Tospovirus	Bunyaviridae	0	1
California encephalitis virus	Orthobunyavirus	Bunyaviridae	0	5
Minireovirus	Norovirus	Caliciviridae	0	1
Newbury-1 virus	Nebovirus	Caliciviridae	0	1
Sapporo virus	Sapovirus	Caliciviridae	39	2744
Norwalk virus	Norovirus	Caliciviridae	9	293
Cassava vein mosaic virus	Cavemovirus	Caulimoviridae	0	1
Rudbeckia flower distortion virus		Caulimoviridae	0	1

Sweetpotato badnavirus A	Badnavirus	Caulimoviridae	0	1
Stilbocarpa mosaic bacilliform virus	Badnavirus	Caulimoviridae	0	10
Pineapple bacilliform virus	Badnavirus	Caulimoviridae	0	24
Taro bacilliform virus	Badnavirus	Caulimoviridae	0	3
Dracaena mottle virus	Badnavirus	Caulimoviridae	0	4
Ambrosia asymptomatic virus 2 UKM-2007	Badnavirus	Caulimoviridae	0	57
Strawberry vein banding virus	Caulimovirus	Caulimoviridae	0	6
Chicken anemia virus	Gyrovirus	Circoviridae	2	15
Beet yellows virus	Closterovirus	Closteroviridae	0	1
Lettuce infectious yellows virus	Crinivirus	Closteroviridae	1	0
Alphacoronavirus 1	Alphacoronavirus	Coronaviridae	0	1
Phytophthora endornavirus 1	Endornavirus	Endornaviridae	1	0
Oryza sativa endornavirus	Endornavirus	Endornaviridae	3	198
Pegivirus A	Pegivirus	Flaviviridae	0	1
Bovine viral diarrhea virus 2	Pestivirus	Flaviviridae	0	36
Bovine viral diarrhea virus 1	Pestivirus	Flaviviridae	0	49
Royal Farm virus	Flavivirus	Flaviviridae	1	0
Dengue virus	Flavivirus	Flaviviridae	1	2
Hepatitis C virus	Hepacivirus	Flaviviridae	1	9
Cotton leaf curl Burewala virus	Begomovirus	Geminiviridae	0	2
Bovine herpesvirus 4	Rhadinovirus	Herpesviridae	0	1
Gallid herpesvirus 1	Iltovirus	Herpesviridae	0	1
Gallid herpesvirus 2	Mardivirus	Herpesviridae	0	1
Gallid herpesvirus 3	Mardivirus	Herpesviridae	0	1
Human herpesvirus 5	Cytomegalovirus	Herpesviridae	0	1
Macacine herpesvirus 1	Simplexvirus	Herpesviridae	0	1
Meleagrid herpesvirus 1	Mardivirus	Herpesviridae	0	1
Murid herpesvirus 2	Muromegalovirus	Herpesviridae	0	1
Ovine herpesvirus 2	Macavirus	Herpesviridae	0	1
Suid herpesvirus 1	Varicellovirus	Herpesviridae	0	1
Ateline herpesvirus 3	Rhadinovirus	Herpesviridae	0	14
Human herpesvirus 3	Varicellovirus	Herpesviridae	0	17
Human herpesvirus 6	Roseolovirus	Herpesviridae	0	19
Bovine herpesvirus 1	Varicellovirus	Herpesviridae	0	2
Canid herpesvirus 1	Varicellovirus	Herpesviridae	0	2
Leporid herpesvirus 4	Simplexvirus	Herpesviridae	0	2
Alcelaphine herpesvirus 1	Macavirus	Herpesviridae	0	20
Macropodid herpesvirus 1	Simplexvirus	Herpesviridae	0	22
Human herpesvirus 8	Rhadinovirus	Herpesviridae	0	25
Equid herpesvirus 1	Varicellovirus	Herpesviridae	0	3
Human herpesvirus 6B	Roseolovirus	Herpesviridae	0	3
Macacine herpesvirus 5	Rhadinovirus	Herpesviridae	0	3
Cercopithecine herpesvirus 9	Varicellovirus	Herpesviridae	0	39
Ateline herpesvirus 2	Rhadinovirus	Herpesviridae	0	4
Saimiriine herpesvirus 2	Rhadinovirus	Herpesviridae	0	43
Anatid herpesvirus 1	Mardivirus	Herpesviridae	0	5
Rodent herpesvirus Peru		Herpesviridae	0	5
Saimiriine herpesvirus 1	Simplexvirus	Herpesviridae	0	9
Bovine herpesvirus 2	Simplexvirus	Herpesviridae	1	0
Bovine herpesvirus 5	Varicellovirus	Herpesviridae	1	0
Equid herpesvirus 4	Varicellovirus	Herpesviridae	1	0
Human herpesvirus 7	Roseolovirus	Herpesviridae	1	0
Equid herpesvirus 9	Varicellovirus	Herpesviridae	1	1
Murid herpesvirus 1	Muromegalovirus	Herpesviridae	1	1
Murid herpesvirus 4	Rhadinovirus	Herpesviridae	1	1
Equid herpesvirus 2	Percavirus	Herpesviridae	1	32
Glossina hytrosavirus	Glossinavirus	Hytrosaviridae	0	1
Musca hytrosavirus	Muscavirus	Hytrosaviridae	0	14
Turbot reddish body iridovirus	Megalocyctivirus	Iridoviridae	0	1
Infectious spleen and kidney necrosis virus	Megalocyctivirus	Iridoviridae	0	12

Invertebrate iridescent virus 9	Iridovirus	Iridoviridae	0	2
Ambystoma tigrinum virus	Ranavirus	Iridoviridae	0	4
Singapore grouper iridovirus	Ranavirus	Iridoviridae	0	43
Bohle iridovirus	Ranavirus	Iridoviridae	0	5
Lymphocystis disease virus - isolate China	Lymphocystivirus	Iridoviridae	0	7
Red sea bream iridovirus	Megalocytivirus	Iridoviridae	0	8
Iridovirus IV31	Iridovirus	Iridoviridae	1	20
Invertebrate iridescent virus 6	Iridovirus	Iridoviridae	1	367
Frog virus 3	Ranavirus	Iridoviridae	1	99
Invertebrate iridescent virus 3	Chloriridovirus	Iridoviridae	2	127
Potato leafroll virus	Polerovirus	Luteoviridae	0	1
Terra1 virus		Mimiviridae	0	150
Acanthamoeba castellanii mamavirus		Mimiviridae	0	23
Courdo11 virus		Mimiviridae	0	6
Acanthamoeba polyphaga mimivirus	Mimivirus	Mimiviridae	11	4328
Moumouvirus		Mimiviridae	1	16
White spot syndrome virus	Whispovirus	Nimaviridae	5	89
Influenza A virus	Influenzavirus A	Orthomyxoviridae	1	85
Alphapapillomavirus 3	Alphapapillomavirus	Papillomaviridae	0	1
Betapapillomavirus 5	Betapapillomavirus	Papillomaviridae	0	1
Gammapapillomavirus 6	Gammapapillomavirus	Papillomaviridae	0	1
Gammapapillomavirus 7	Gammapapillomavirus	Papillomaviridae	0	1
Pudu puda papillomavirus		Papillomaviridae	0	1
Gammapapillomavirus 5	Gammapapillomavirus	Papillomaviridae	0	10
Gammapapillomavirus 2	Gammapapillomavirus	Papillomaviridae	0	2
Macaca fascicularis papillomavirus	Betapapillomavirus	Papillomaviridae	0	3
Gammapapillomavirus 4	Gammapapillomavirus	Papillomaviridae	0	4
Human papillomavirus 132-like viruses	Gammapapillomavirus	Papillomaviridae	0	5
Alphapapillomavirus 9	Alphapapillomavirus	Papillomaviridae	1	0
Gammapapillomavirus 1	Gammapapillomavirus	Papillomaviridae	1	8
Bovine respiratory syncytial virus	Pneumovirus	Paramyxoviridae	0	19
Ovine respiratory syncytial virus	Pneumovirus	Paramyxoviridae	0	2
Respiratory syncytial virus	Pneumovirus	Paramyxoviridae	2	25
Human respiratory syncytial virus	Pneumovirus	Paramyxoviridae	7	377
Beet cryptic virus 2	Alphacryptovirus	Partitiviridae	0	13
Cryptosporidium parvum virus 1	Cryspovirus	Partitiviridae	0	4
Penicillium stoloniferum virus F	Partitivirus	Partitiviridae	0	5
Adeno-associated virus 11	Dependovirus	Parvoviridae	0	1
Adeno-associated virus 12	Dependovirus	Parvoviridae	0	1
Adeno-associated virus - 1	Dependovirus	Parvoviridae	0	1
Adeno-associated virus - 7	Dependovirus	Parvoviridae	0	1
Adeno-associated virus - 8	Dependovirus	Parvoviridae	0	1
Adeno-associated virus - Po2	Dependovirus	Parvoviridae	0	1
Aplomado falcon AAV2	Dependovirus	Parvoviridae	0	1
Blattella germanica densovirus	Pefudensovirus	Parvoviridae	0	1
Duck parvovirus	Dependovirus	Parvoviridae	0	1
Goose parvovirus	Dependovirus	Parvoviridae	0	1
Rat adeno-associated virus 1	Dependovirus	Parvoviridae	0	1
Adeno-associated virus - 3	Dependovirus	Parvoviridae	0	11
Non-human primate Adeno-associated virus	Dependovirus	Parvoviridae	0	15
Adeno-associated virus 9	Dependovirus	Parvoviridae	0	2
Adeno-associated virus VR-195	Dependovirus	Parvoviridae	0	2
Aedes albopictus densovirus	Brevidensovirus	Parvoviridae	0	2
Adeno-associated virus 13	Dependovirus	Parvoviridae	0	3
Adeno-associated virus VR-355	Dependovirus	Parvoviridae	0	3
Adeno-associated virus 10	Dependovirus	Parvoviridae	0	4
Adeno-associated virus - 2	Dependovirus	Parvoviridae	0	80
Planococcus citri densovirus	Pefudensovirus	Parvoviridae	1	22
Acheta domestica densovirus	Pefudensovirus	Parvoviridae	1	4
Adeno-associated virus	Dependovirus	Parvoviridae	15	465

Pyramimonas orientalis virus		Phycodnaviridae	0	1
Paramecium bursaria Chlorella virus				
Marburg 1	Chlorovirus	Phycodnaviridae	0	10
Feldmannia irregularis virus a	Phaeovirus	Phycodnaviridae	0	11
Paramecium bursaria Chlorella virus				
NYS1	Chlorovirus	Phycodnaviridae	0	13
Paramecium bursaria Chlorella virus				
NC1A	Chlorovirus	Phycodnaviridae	0	5
Ostreococcus lucimarinus virus OIV1	Prasinovirus	Phycodnaviridae	0	79
Ostreococcus tauri virus 1		Phycodnaviridae	0	81
Emiliana huxleyi virus 86	Coccolithovirus	Phycodnaviridae	0	88
Micromonas sp. RCC1109 virus MpV1	Prasinovirus	Phycodnaviridae	1	122
Chlorella virus	Chlorovirus	Phycodnaviridae	1	18
Paramecium bursaria Chlorella virus A1	Chlorovirus	Phycodnaviridae	12	367
Ostreococcus tauri virus 2		Phycodnaviridae	1	348
Paramecium bursaria Chlorella virus				
NY2A	Chlorovirus	Phycodnaviridae	2	156
Paramecium bursaria Chlorella virus 1	Chlorovirus	Phycodnaviridae	3	118
Feldmannia species virus	Phaeovirus	Phycodnaviridae	3	17
Ostreococcus virus OsV5		Phycodnaviridae	3	24
Ectocarpus siliculosus virus 1	Phaeovirus	Phycodnaviridae	4	364
Bathycoccus sp. RCC1105 virus BpV	Prasinovirus	Phycodnaviridae	4	768
Emiliana huxleyi virus 99B1	Coccolithovirus	Phycodnaviridae	5	160
Paramecium bursaria Chlorella virus				
AR158	Chlorovirus	Phycodnaviridae	5	241
Acanthocystis turfacea Chlorella virus 1	Chlorovirus	Phycodnaviridae	8	330
Human picobirnavirus	Picobirnavirus	Picobirnaviridae	1	8
Encephalomyocarditis virus	Cardiovirus	Picornaviridae	0	1
Enterovirus A	Enterovirus	Picornaviridae	0	1
Human rhinovirus hRVLz148	Enterovirus	Picornaviridae	0	1
Duck hepatitis A virus	Avihepatovirus	Picornaviridae	1	0
Rhinovirus A	Enterovirus	Picornaviridae	1	0
Enterovirus B	Enterovirus	Picornaviridae	1	157
Human parechovirus	Parechovirus	Picornaviridae	27	8603
Rhinovirus C	Enterovirus	Picornaviridae	3	50
Hyposoter fugitivus ichnovirus	Ichnovirus	Polydnaviridae	0	3
Glypta fumiferanae ichnovirus	Ichnovirus	Polydnaviridae	0	33
Cotesia congregata bracovirus	Bracovirus	Polydnaviridae	0	6
Merkel cell polyomavirus	Polyomavirus	Polyomaviridae	0	1
Simian virus 40	Polyomavirus	Polyomaviridae	0	1
Chimpanzee polyomavirus	Polyomavirus	Polyomaviridae	1	0
Turnip mosaic virus	Potyvirus	Potyviridae	0	1
Cassava brown streak virus	Ipomovirus	Potyviridae	0	12
Iris mild mosaic virus	Potyvirus	Potyviridae	0	2
Wheat streak mosaic virus	Tritimovirus	Potyviridae	0	2
Papaya ringspot virus	Potyvirus	Potyviridae	0	22
Cypovirus 1	Cypovirus	Reoviridae	0	1
Murine rotavirus	Rotavirus	Reoviridae	0	1
Rice ragged stunt virus	Oryzavirus	Reoviridae	0	1
Great Island virus	Orbivirus	Reoviridae	0	10
Equine rotavirus	Rotavirus	Reoviridae	0	2
Mammalian orthoreovirus	Orthoreovirus	Reoviridae	0	2
Rotavirus C	Rotavirus	Reoviridae	0	2
Feline rotavirus	Rotavirus	Reoviridae	0	3
Rotavirus sp.	Rotavirus	Reoviridae	0	3
Bluetongue virus	Orbivirus	Reoviridae	1	7
Rotavirus A	Rotavirus	Reoviridae	83	996
Avian retrovirus IC10	Alpharetrovirus	Retroviridae	0	1
Feline foamy virus	Spumavirus	Retroviridae	0	1
Mason-Pfizer monkey virus	Betaretrovirus	Retroviridae	0	1
Murine leukemia-related retroviruses	Gammaretrovirus	Retroviridae	0	1
Rous sarcoma virus	Alpharetrovirus	Retroviridae	0	1

Simian foamy virus	Spumavirus	Retroviridae	0	1
Simian immunodeficiency virus	Lentivirus	Retroviridae	0	1
Human endogenous retrovirus K		Retroviridae	0	12
Human endogenous retrovirus		Retroviridae	0	2
Human endogenous retrovirus W		Retroviridae	0	22
Feline leukemia virus	Gammaretrovirus	Retroviridae	0	34
Spleen focus-forming virus	Gammaretrovirus	Retroviridae	0	4
Human endogenous retrovirus H		Retroviridae	0	56
AKT8 retrovirus		Retroviridae	0	58
Murine leukemia virus	Gammaretrovirus	Retroviridae	0	6
Snakehead retrovirus		Retroviridae	1	0
Feline immunodeficiency virus	Lentivirus	Retroviridae	1	2
Human immunodeficiency virus 1	Lentivirus	Retroviridae	3	116
Wheat rosette stunt virus		Rhabdoviridae	0	3
Maize chlorotic dwarf virus	Waikavirus	Secoviridae	1	0
Saccharomyces cerevisiae virus L-A	Totivirus	Totiviridae	0	1
Ononis yellow mosaic virus	Tymovirus	Tymoviridae	0	1
Rehmannia mosaic virus	Tobamovirus	Virgaviridae	0	1
Tobacco mosaic virus	Tobamovirus	Virgaviridae	0	3
Pepper mild mottle virus	Tobamovirus	Virgaviridae	1	54

Supplemental Table S16. Viral output of translated nucleotide alignment to viral protein DB for a prostate cancer NGS dataset, containing HPV-18.

Species	Genus	Family	Reads #80
Alphapapillomavirus 7	Alphapapillomavirus	Papillomaviridae	2

Supplemental Table S17. Viral output of translated nucleotide alignment to viral protein DB for an encephalitis NGS dataset (G), containing HHV-3.

Species	Genus	Family	Reads #0
Circovirus-like genome RW-E			1
Abalone shriveling syndrome-associated virus			44
Heliothis zea virus 1			77
Oryctes rhinoceros virus			8
Porcine adenovirus C	Mastadenovirus	Adenoviridae	1
Anguillid herpesvirus 1	Cyprinivirus	Alloherpesviridae	1
Ictalurid herpesvirus 1	Ictalurivirus	Alloherpesviridae	1
Ranid herpesvirus 2	Batrachovirus	Alloherpesviridae	1
Cyprinid herpesvirus 3	Cyprinivirus	Alloherpesviridae	2
Salmonid herpesvirus 2	Salmonivirus	Alloherpesviridae	2
SEN virus		Anelloviridae	1
Torque teno virus 12	Alphatorquevirus	Anelloviridae	1
TTV-like mini virus		Anelloviridae	1
Torque teno midi virus		Anelloviridae	2
Torque teno virus 11	Alphatorquevirus	Anelloviridae	2
Torque teno virus 13	Alphatorquevirus	Anelloviridae	2
Torque teno virus 2	Alphatorquevirus	Anelloviridae	2
Torque teno mini virus 2	Betatorquevirus	Anelloviridae	3
Torque teno virus		Anelloviridae	6
Spodoptera frugiperda ascovirus 1a	Ascovirus	Ascoviridae	1
Trichoplusia ni ascovirus 2c	Ascovirus	Ascoviridae	16
Diadromus pulchellus ascovirus 4a	Ascovirus	Ascoviridae	32
Heliothis virescens ascovirus 3a	Ascovirus	Ascoviridae	3
African swine fever virus	Asfivirus	Asfarviridae	8
Avastrovirus 2	Avastrovirus	Astroviridae	1
Mamastrovirus 1	Mamastrovirus	Astroviridae	8
Culex nigripalpus nucleopolyhedrovirus	Deltabaculovirus	Baculoviridae	1
Ectropis obliqua nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	1
Helicoverpa armigera granulovirus	Betabaculovirus	Baculoviridae	1
Leucania separata nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	1
Lymantria dispar multiple nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	1
Pseudaletia unipuncta granulovirus	Betabaculovirus	Baculoviridae	1
Spodoptera exigua multiple nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	1
Spodoptera litura nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	1
Antheraea pernyi nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	2
Cydia pomonella granulovirus	Betabaculovirus	Baculoviridae	2
Hyphantria cunea nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	2
Orgyia ericae nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	2
Orgyia leucostigma NPV	Alphabaculovirus	Baculoviridae	2
Orgyia pseudotsugata multiple nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	2
Spodoptera litura nucleopolyhedrovirus II	Alphabaculovirus	Baculoviridae	2
Agrotis segetum nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	3
Trichoplusia ni single nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	3
Autographa californica multiple nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	4
Neodiprion lecontei nucleopolyhedrovirus	Gammabaculovirus	Baculoviridae	4
Chrysodeixis chalcites nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	5
Euproctis pseudoconspersa nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	5
Lymantria xylini MNPV	Alphabaculovirus	Baculoviridae	5
St-Valerien swine virus		Caliciviridae	1
Norwalk virus	Norovirus	Caliciviridae	22
Sweet potato vein clearing virus	Solendovirus	Caulimoviridae	1
Ambrosia asymptomatic virus 2 UKM-2007	Badnavirus	Caulimoviridae	2
Cacao swollen shoot virus	Badnavirus	Caulimoviridae	2
Gull circovirus	Circovirus	Circoviridae	2
Citrus tristeza virus	Closterovirus	Closteroviridae	1

Little cherry virus 1		Closteroviridae	1
Plum bark necrosis stem pitting-associated virus	Ampelovirus	Closteroviridae	2
Hepatitis C virus	Hepacivirus	Flaviviridae	10
Bovine viral diarrhea virus 2	Pestivirus	Flaviviridae	1
Bovine viral diarrhea virus 1	Pestivirus	Flaviviridae	3
East African cassava mosaic virus	Begomovirus	Geminiviridae	1
Hepatitis B virus	Orthohepadnavirus	Hepadnaviridae	2
Anatid herpesvirus 1	Mardivirus	Herpesviridae	1
Bovine herpesvirus 5	Varicellovirus	Herpesviridae	1
Gallid herpesvirus 1	Iltovirus	Herpesviridae	1
Macropodid herpesvirus 1	Simplexvirus	Herpesviridae	1
Meleagrid herpesvirus 1	Mardivirus	Herpesviridae	1
Papiine herpesvirus 2	Simplexvirus	Herpesviridae	1
Ateline herpesvirus 2	Rhadinovirus	Herpesviridae	2
Bovine herpesvirus 1	Varicellovirus	Herpesviridae	2
Equid herpesvirus 1	Varicellovirus	Herpesviridae	2
Rodent herpesvirus Peru		Herpesviridae	2
Saimiriine herpesvirus 2	Rhadinovirus	Herpesviridae	2
Tupaiaid herpesvirus 1		Herpesviridae	2
Cercopithecine herpesvirus 9	Varicellovirus	Herpesviridae	25
Caviid herpesvirus 2		Herpesviridae	3
Cercopithecine herpesvirus 2	Simplexvirus	Herpesviridae	3
Murid herpesvirus 4	Rhadinovirus	Herpesviridae	4
Felid herpesvirus 1	Varicellovirus	Herpesviridae	4
Human herpesvirus 1	Simplexvirus	Herpesviridae	4
Human herpesvirus 8	Rhadinovirus	Herpesviridae	4
Ovine herpesvirus 2	Macavirus	Herpesviridae	5
Saimiriine herpesvirus 1	Simplexvirus	Herpesviridae	5
Macacine herpesvirus 5	Rhadinovirus	Herpesviridae	6
Ateline herpesvirus 3	Rhadinovirus	Herpesviridae	7
Human herpesvirus 6	Roseolovirus	Herpesviridae	7
Suid herpesvirus 1	Varicellovirus	Herpesviridae	7
Human herpesvirus 3	Varicellovirus	Herpesviridae	8
Equid herpesvirus 2	Percavirus	Herpesviridae	9
Musca hytrosavirus	Muscavirus	Hytrosaviridae	7
Frog virus 3	Ranavirus	Iridoviridae	10
Invertebrate iridescent virus 16	Iridovirus	Iridoviridae	1
Iridovirus IV31	Iridovirus	Iridoviridae	1
Red sea bream iridovirus	Megalocyctivirus	Iridoviridae	1
Turbot reddish body iridovirus	Megalocyctivirus	Iridoviridae	1
Acetes erythraeus iridovirus	Iridovirus	Iridoviridae	2
Ambystoma tigrinum virus	Ranavirus	Iridoviridae	2
Lymphocystis disease virus 1	Lymphocystivirus	Iridoviridae	3
Epizootic haematopoietic necrosis virus	Ranavirus	Iridoviridae	4
Singapore grouper iridovirus	Ranavirus	Iridoviridae	4
Lymphocystis disease virus - isolate China	Lymphocystivirus	Iridoviridae	47
Invertebrate iridescent virus 6	Iridovirus	Iridoviridae	51
Invertebrate iridescent virus 3	Chloriridovirus	Iridoviridae	60
Courdo11 virus		Mimiviridae	12
Acanthamoeba castellanii mamavirus		Mimiviridae	2
Moumouvirus		Mimiviridae	2
Acanthamoeba polyphaga mimivirus	Mimivirus	Mimiviridae	307
Terra1 virus		Mimiviridae	3
White spot syndrome virus	Whispovirus	Nimaviridae	16
Paramecium bursaria Chlorella virus AR158	Chlorovirus	Phycodnaviridae	108
Chrysochromulina ericina virus		Phycodnaviridae	1
Phaeocystis globosa virus	Prymnesiovirus	Phycodnaviridae	1
Chlorella virus	Chlorovirus	Phycodnaviridae	14
Feldmannia irregularis virus a	Phaeovirus	Phycodnaviridae	14
Feldmannia species virus	Phaeovirus	Phycodnaviridae	18
Pyramimonas orientalis virus		Phycodnaviridae	2
Paramecium bursaria Chlorella virus A1	Chlorovirus	Phycodnaviridae	227

Ostreococcus tauri virus 1		Phycodnaviridae	23
Bathycoccus sp. RCC1105 virus BpV	Prasinovirus	Phycodnaviridae	237
Acanthocystis turfacea Chlorella virus 1	Chlorovirus	Phycodnaviridae	261
Emiliana huxleyi virus 99B1	Coccolithovirus	Phycodnaviridae	34
Paramecium bursaria Chlorella virus NY2A	Chlorovirus	Phycodnaviridae	46
Emiliana huxleyi virus 86	Coccolithovirus	Phycodnaviridae	49
Ostreococcus lucimarinus virus OIV1	Prasinovirus	Phycodnaviridae	62
Ectocarpus siliculosus virus 1	Phaeovirus	Phycodnaviridae	67
Ostreococcus virus OsV5		Phycodnaviridae	67
Ostreococcus tauri virus 2		Phycodnaviridae	69
Micromonas sp. RCC1109 virus MpV1	Prasinovirus	Phycodnaviridae	87
Paramecium bursaria Chlorella virus 1	Chlorovirus	Phycodnaviridae	95
Cosavirus A	Cosavirus	Picornaviridae	1
Foot-and-mouth disease virus	Aphthovirus	Picornaviridae	1
Enterovirus A	Enterovirus	Picornaviridae	4
Glypta fumiferanae ichnovirus	Ichnovirus	Polydnaviridae	9
Chimpanzee polyomavirus	Polyomavirus	Polyomaviridae	1
Goose hemorrhagic polyomavirus	Polyomavirus	Polyomaviridae	1
Polyomavirus HPyV6	Polyomavirus	Polyomaviridae	1
Finch polyomavirus	Polyomavirus	Polyomaviridae	3
Sorghum mosaic virus	Potyvirus	Potyviridae	1
Ugandan cassava brown streak virus	Ipomovirus	Potyviridae	1
Cassava brown streak virus	Ipomovirus	Potyviridae	6
Iris mild mosaic virus	Potyvirus	Potyviridae	7
Rotavirus C	Rotavirus	Reoviridae	5
Human immunodeficiency virus 1	Lentivirus	Retroviridae	1
Atlantic salmon swim bladder sarcoma virus		Retroviridae	1
Avian carcinoma Mill Hill virus 2	Alpharetrovirus	Retroviridae	1
Avian retrovirus IC10	Alpharetrovirus	Retroviridae	1
Avian sarcoma virus	Alpharetrovirus	Retroviridae	1
Baboon endogenous virus	Gammaretrovirus	Retroviridae	1
Feline leukemia virus	Gammaretrovirus	Retroviridae	1
Gibbon ape leukemia virus	Gammaretrovirus	Retroviridae	1
Murine leukemia virus	Gammaretrovirus	Retroviridae	1
Puma lentivirus	Lentivirus	Retroviridae	1
Human endogenous retrovirus H		Retroviridae	15
LNras*SN acutely transforming retrovirus		Retroviridae	2
Avian leukosis virus	Alpharetrovirus	Retroviridae	3
Porcine type-C oncovirus	Gammaretrovirus	Retroviridae	3
Human endogenous retrovirus		Retroviridae	4
Human endogenous retrovirus W		Retroviridae	7
Human endogenous retrovirus K		Retroviridae	8

Supplemental Table S18. Viral output of translated nucleotide alignment to viral protein DB for a TTVS hepatitis NGS dataset (E), containing HCV-1b.

Species	Genus	Family	Contigs #CACGAT	Reads #CACGAT
Abalone shriveling syndrome-associated virus			0	1
Pyrococcus abyssi virus 1			1	0
Heliothis zea virus 1			1	15
Anguillid herpesvirus 1	Cyprinivirus	Alloherpesviridae	0	11
Pepino mosaic virus	Potexvirus	Alphaflexiviridae	0	1
Shallot virus X	Allexivirus	Alphaflexiviridae	0	1
SEN virus		Anelloviridae	0	1
Torque teno midi virus 2	Gammatorquevirus	Anelloviridae	0	1
Torque teno virus 14	Alphatorquevirus	Anelloviridae	0	1
Torque teno virus 19	Alphatorquevirus	Anelloviridae	0	1
Torque teno virus 23	Alphatorquevirus	Anelloviridae	0	1
Torque teno virus 24	Alphatorquevirus	Anelloviridae	0	1
Torque teno mini virus 1	Betatorquevirus	Anelloviridae	0	2
Torque teno virus 2	Alphatorquevirus	Anelloviridae	0	25
Torque teno mini virus 2	Betatorquevirus	Anelloviridae	0	5
Torque teno virus 10	Alphatorquevirus	Anelloviridae	1	1
TTV-like mini virus		Anelloviridae	1	1
Torque teno virus 3	Alphatorquevirus	Anelloviridae	1	2
Torque teno virus 13	Alphatorquevirus	Anelloviridae	1	30
Torque teno virus 16	Alphatorquevirus	Anelloviridae	1	4
Torque teno virus		Anelloviridae	5	162
Trichoplusia ni ascovirus 2c	Ascovirus	Ascoviridae	1	5
Diadromus pulchellus ascovirus 4a	Ascovirus	Ascoviridae	1	9
African swine fever virus	Asfivirus	Asfarviridae	0	1
Agrotis segetum nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	0	1
Spodoptera litura nucleopolyhedrovirus II	Alphabaculovirus	Baculoviridae	0	1
Epinotia aporema granulovirus	Betabaculovirus	Baculoviridae	0	2
Pieris rapae granulovirus	Betabaculovirus	Baculoviridae	0	2
Spodoptera litura nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	0	28
Ectopis obliqua nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	0	3
Autographa californica multiple nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	2	12
Helleborus net necrosis virus	Carlavirus	Betaflexiviridae	0	1
Sandfly fever Naples virus	Phlebovirus	Bunyaviridae	1	0
Little cherry virus 1		Closteroviridae	0	1
Avian coronavirus	Gammacoronavirus	Coronaviridae	0	1
Pegivirus B	Pegivirus	Flaviviridae	0	1
Bovine viral diarrhea virus 2	Pestivirus	Flaviviridae	0	16
Pegivirus A	Pegivirus	Flaviviridae	0	18
Bovine viral diarrhea virus 1	Pestivirus	Flaviviridae	0	3
Hepatitis GB virus C-like virus	Pegivirus	Flaviviridae	0	3
GB virus C	Pegivirus	Flaviviridae	110	37409
Hepatitis C virus	Hepacivirus	Flaviviridae	41	4647
Hepatitis B virus	Orthohepadnavirus	Hepadnaviridae	0	2
Callitrichine herpesvirus 3	Lymphocryptovirus	Herpesviridae	0	1
Equid herpesvirus 2	Percavirus	Herpesviridae	0	1
Equid herpesvirus 4	Varicellovirus	Herpesviridae	0	1
Felid herpesvirus 1	Varicellovirus	Herpesviridae	0	1
Human herpesvirus 3	Varicellovirus	Herpesviridae	0	1
Human herpesvirus 6	Roseolovirus	Herpesviridae	0	1
Rodent herpesvirus Peru		Herpesviridae	0	1
Saimiriine herpesvirus 2	Rhadinovirus	Herpesviridae	0	1
Suiformes rhadinovirus genogroup 1	Rhadinovirus	Herpesviridae	0	1
Bovine herpesvirus 1	Varicellovirus	Herpesviridae	0	13

Human herpesvirus 8	Rhadinovirus	Herpesviridae	0	2
Caviid herpesvirus 2		Herpesviridae	0	3
Tupaïid herpesvirus 1		Herpesviridae	0	3
Suid herpesvirus 1	Varicellovirus	Herpesviridae	0	32
Cercopithecine herpesvirus 9	Varicellovirus	Herpesviridae	0	4
Human herpesvirus 7	Roseolovirus	Herpesviridae	0	5
Ateline herpesvirus 3	Rhadinovirus	Herpesviridae	1	0
Bovine herpesvirus 4	Rhadinovirus	Herpesviridae	1	0
Gallid herpesvirus 2	Mardivirus	Herpesviridae	1	0
Ovine herpesvirus 2	Macavirus	Herpesviridae	1	0
Equid herpesvirus 9	Varicellovirus	Herpesviridae	1	2
Murid herpesvirus 4	Rhadinovirus	Herpesviridae	1	4
Glossina hytivirus	Glossinavirus	Hytrosaviridae	0	1
Musca hytivirus	Muscavirus	Hytrosaviridae	0	1
Red sea bream iridovirus	Megalocytiavirus	Iridoviridae	0	1
Invertebrate iridescent virus 6	Iridovirus	Iridoviridae	0	32
Singapore grouper iridovirus	Ranavirus	Iridoviridae	0	7
Courdo11 virus		Mimiviridae	1	13
Acanthamoeba polyphaga mimivirus	Mimivirus	Mimiviridae	3	36
White spot syndrome virus	Whispovirus	Nimaviridae	0	3
Human respiratory syncytial virus	Pneumovirus	Paramyxoviridae	1	0
Human parvovirus 4	Parvovirus	Parvoviridae	1	0
Ostreococcus virus OsV5		Phycodnaviridae	0	1
Paramecium bursaria Chlorella virus NY2A	Chlorovirus	Phycodnaviridae	0	15
Feldmannia species virus	Phaeovirus	Phycodnaviridae	0	2
Micromonas sp. RCC1109 virus MpV1	Prasinovirus	Phycodnaviridae	0	2
Ostreococcus lucimarinus virus OIV1	Prasinovirus	Phycodnaviridae	0	2
Emiliana huxleyi virus 99B1	Coccolithovirus	Phycodnaviridae	0	3
Chlorella virus	Chlorovirus	Phycodnaviridae	1	2
Bathycoccus sp. RCC1105 virus BpV	Prasinovirus	Phycodnaviridae	1	22
Paramecium bursaria Chlorella virus 1	Chlorovirus	Phycodnaviridae	1	29
Paramecium bursaria Chlorella virus AR158	Chlorovirus	Phycodnaviridae	1	5
Ostreococcus tauri virus 2		Phycodnaviridae	1	8
Emiliana huxleyi virus 86	Coccolithovirus	Phycodnaviridae	3	21
Acanthocystis turfacea Chlorella virus 1	Chlorovirus	Phycodnaviridae	3	33
Ectocarpus siliculosus virus 1	Phaeovirus	Phycodnaviridae	4	35
Paramecium bursaria Chlorella virus A1	Chlorovirus	Phycodnaviridae	5	41
Encephalomyocarditis virus	Cardiovirus	Picornaviridae	0	2
Cotesia congregata bracovirus	Bracovirus	Polydnaviridae	0	1
Hyposoter fugitivus ichnovirus	Ichnovirus	Polydnaviridae	0	8
Canary polyomavirus	Polyomavirus	Polyomaviridae	0	2
Iris mild mosaic virus	Potyvirus	Potyviridae	1	6
Human endogenous retrovirus		Retroviridae	0	1
Murine osteosarcoma virus	Gammaretrovirus	Retroviridae	0	1
Rabbit endogenous retrovirus H	Betaretrovirus	Retroviridae	0	1
Feline leukemia virus	Gammaretrovirus	Retroviridae	0	13
Human endogenous retrovirus H		Retroviridae	0	14
LNras*SN acutely transforming retrovirus		Retroviridae	0	4
Human endogenous retrovirus K		Retroviridae	0	5
Human immunodeficiency virus 1	Lentivirus	Retroviridae	2	3
Lyssavirus Ozernoe	Lyssavirus	Rhabdoviridae	0	5
Phlebiopsis gigantea mycovirus dsRNA 1		Totiviridae	1	0

Supplemental Table S19. Viral output of translated nucleotide alignment to viral protein DB for a simian pneumonia NGS dataset (K), containing TMAdV.

Species	Genus	Family	Contigs #0	Reads #0
Bat adenovirus FBV1		Adenoviridae	0	1
Bovine adenovirus A	Mastadenovirus	Adenoviridae	0	1
Falcon adenovirus A	Aviadenovirus	Adenoviridae	0	1
Great tit adenovirus A	Siadenovirus	Adenoviridae	0	1
Sturgeon adenovirus A	Ichtadenovirus	Adenoviridae	0	1
Ovine adenovirus A	Mastadenovirus	Adenoviridae	0	10
Equine adenovirus B	Mastadenovirus	Adenoviridae	0	14
Murine adenovirus B	Mastadenovirus	Adenoviridae	0	14
Ovine adenovirus B	Mastadenovirus	Adenoviridae	0	14
Bovine adenovirus C	Mastadenovirus	Adenoviridae	0	15
Murine adenovirus C	Mastadenovirus	Adenoviridae	0	2
unidentified adenovirus	Mastadenovirus	Adenoviridae	0	2
Caprine adenovirus 2	Mastadenovirus	Adenoviridae	0	3
Porcine adenovirus C	Mastadenovirus	Adenoviridae	0	40
Bat adenovirus A	Mastadenovirus	Adenoviridae	0	6
Cynomolgus adenovirus 1	Mastadenovirus	Adenoviridae	0	6
Porcine adenovirus A	Mastadenovirus	Adenoviridae	0	8
Tree shrew adenovirus A	Mastadenovirus	Adenoviridae	0	9
Human adenovirus G	Mastadenovirus	Adenoviridae	1	22
Canine adenovirus A	Mastadenovirus	Adenoviridae	1	23
Simian adenovirus B	Mastadenovirus	Adenoviridae	1	29
Simian adenovirus A	Mastadenovirus	Adenoviridae	13	177
Simian adenovirus 7	Mastadenovirus	Adenoviridae	2	12
Bovine adenovirus B	Mastadenovirus	Adenoviridae	2	30
Human adenovirus C	Mastadenovirus	Adenoviridae	2	49
Human adenovirus F	Mastadenovirus	Adenoviridae	3	60
Human adenovirus D	Mastadenovirus	Adenoviridae	3	67
Human adenovirus A	Mastadenovirus	Adenoviridae	3	80
Human adenovirus B	Mastadenovirus	Adenoviridae	4	175
Human adenovirus E	Mastadenovirus	Adenoviridae	5	77
Pepino mosaic virus	Potexvirus	Alphaflexiviridae	0	8
Maruca vitrata nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	0	1
Autographa californica multiple nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	0	4
Columbid circovirus	Circovirus	Circoviridae	0	5
Border disease virus	Pestivirus	Flaviviridae	0	1
Bovine viral diarrhea virus 2	Pestivirus	Flaviviridae	0	11
Bovine viral diarrhea virus 1	Pestivirus	Flaviviridae	0	4
Human herpesvirus 6	Roseolovirus	Herpesviridae	0	1
Saimiriine herpesvirus 2	Rhadinovirus	Herpesviridae	0	1
Lymphocystis disease virus - isolate China	Lymphocystivirus	Iridoviridae	1	1
Influenza A virus	Influenzavirus A	Orthomyxoviridae	0	1
Alphapapillomavirus 7	Alphapapillomavirus	Papillomaviridae	0	4
Acanthocystis turfacea Chlorella virus 1	Chlorovirus	Phycodnaviridae	0	1
Feldmannia species virus	Phaeovirus	Phycodnaviridae	0	1
Emiliana huxleyi virus 86	Coccolithovirus	Phycodnaviridae	0	2
Simian enterovirus SV46	Enterovirus	Picornaviridae	0	1
Simian picornavirus 6		Picornaviridae	0	1
Enterovirus J	Enterovirus	Picornaviridae	0	15
Simian picornavirus 8		Picornaviridae	0	2
Enterovirus A	Enterovirus	Picornaviridae	0	61
Glypta fumiferanae ichtnovirus	Ichnovirus	Polydnaviridae	1	10
Rotavirus A	Rotavirus	Reoviridae	1	9

AKT8 retrovirus		Retroviridae	0	1
Avian carcinoma Mill Hill virus 2	Alpharetrovirus	Retroviridae	0	1
Avian musculoaponeurotic fibrosarcoma virus AS42	Alpharetrovirus	Retroviridae	0	1
Hardy-Zuckerman feline sarcoma virus	Gammaretrovirus	Retroviridae	0	1
Mus musculus mobilized endogenous polytropic provirus	Gammaretrovirus	Retroviridae	0	1
Odokoileus hemionus endogenous retrovirus	Gammaretrovirus	Retroviridae	0	1
Rabbit endogenous retrovirus H	Betaretrovirus	Retroviridae	0	1
Recombinant M-MuLV RaLV retrovirus	Gammaretrovirus	Retroviridae	0	1
Woolly monkey sarcoma virus	Gammaretrovirus	Retroviridae	0	1
Baboon endogenous virus	Gammaretrovirus	Retroviridae	0	10
Enzootic nasal tumour virus of goats	Betaretrovirus	Retroviridae	0	16
Feline leukemia virus	Gammaretrovirus	Retroviridae	0	18
Simian retrovirus 4	Betaretrovirus	Retroviridae	0	19
Avian leukosis virus	Alpharetrovirus	Retroviridae	0	2
Human mammary tumor virus	Betaretrovirus	Retroviridae	0	2
Moloney murine sarcoma virus	Gammaretrovirus	Retroviridae	0	2
Human endogenous retrovirus K		Retroviridae	0	27
Python molurus endogenous retrovirus		Retroviridae	0	3
Murine leukemia virus	Gammaretrovirus	Retroviridae	0	33
Atlantic salmon swim bladder sarcoma virus		Retroviridae	0	4
Porcine type-C oncovirus	Gammaretrovirus	Retroviridae	0	44
Mouse mammary tumor virus	Betaretrovirus	Retroviridae	0	5
Gibbon ape leukemia virus	Gammaretrovirus	Retroviridae	0	9
Murine leukemia-related retroviruses	Gammaretrovirus	Retroviridae	0	9
Human endogenous retrovirus H		Retroviridae	1	116
RD114 retrovirus	Gammaretrovirus	Retroviridae	1	20
Ovine enzootic nasal tumor virus	Betaretrovirus	Retroviridae	1	24
Jaagsiekte sheep retrovirus	Betaretrovirus	Retroviridae	1	6
Mason-Pfizer monkey virus	Betaretrovirus	Retroviridae	2	36
Squirrel monkey retrovirus	Betaretrovirus	Retroviridae	2	7
Human endogenous retrovirus		Retroviridae	3	139
Human endogenous retrovirus W		Retroviridae	4	123
Reticuloendotheliosis virus	Gammaretrovirus	Retroviridae	5	57

Supplemental Table S20. Viral output of translated nucleotide alignment for a hemorrhagic fever NGS dataset (D), containing BASV.

Species	Genus	Family	Contigs #9	Reads #9
Heliothis zea virus 1			0	3
Porcine reproductive and respiratory syndrome virus	Arterivirus	Arteriviridae	1	0
Diadromus pulchellus ascovirus 4a	Ascovirus	Ascoviridae	0	2
Trichoplusia ni ascovirus 2c	Ascovirus	Ascoviridae	0	4
African swine fever virus	Asfivirus	Asfarviridae	0	1
Agrotis segetum granulovirus	Betabaculovirus	Baculoviridae	0	1
Choristoneura occidentalis granulovirus	Betabaculovirus	Baculoviridae	1	0
Sapporo virus	Sapovirus	Caliciviridae	1	0
Bovine viral diarrhea virus 2	Pestivirus	Flaviviridae	0	1
Macacine herpesvirus 5	Rhadinovirus	Herpesviridae	0	18
Equid herpesvirus 2	Percavirus	Herpesviridae	0	2
Saimiriine herpesvirus 2	Rhadinovirus	Herpesviridae	0	4
Pan troglodytes	Pan	Hominidae	0	1
Infectious spleen and kidney necrosis virus	Megalocytivirus	Iridoviridae	0	1
Invertebrate iridescent virus 3	Chloriridovirus	Iridoviridae	0	1
Acanthamoeba polyphaga mimivirus	Mimivirus	Mimiviridae	0	18
White spot syndrome virus	Whispovirus	Nimaviridae	0	1
Emiliana huxleyi virus 86	Coccolithovirus	Phycodnaviridae	0	1
Ostreococcus virus OsV5		Phycodnaviridae	0	2
Feldmannia species virus	Phaeovirus	Phycodnaviridae	0	4
Paramecium bursaria Chlorella virus 1	Chlorovirus	Phycodnaviridae	0	4
Paramecium bursaria Chlorella virus AR158	Chlorovirus	Phycodnaviridae	0	4
Micromonas sp. RCC1109 virus MpV1	Prasinovirus	Phycodnaviridae	0	5
Bathycoccus sp. RCC1105 virus BpV	Prasinovirus	Phycodnaviridae	0	8
Feldmannia irregularis virus a	Phaeovirus	Phycodnaviridae	1	0
Acanthocystis turfacea Chlorella virus 1	Chlorovirus	Phycodnaviridae	1	24
Paramecium bursaria Chlorella virus A1	Chlorovirus	Phycodnaviridae	1	4
Ectocarpus siliculosus virus 1	Phaeovirus	Phycodnaviridae	1	5
Cotesia congregata bracovirus	Bracovirus	Polydnaviridae	0	1
Cotesia glomerata bracovirus	Bracovirus	Polydnaviridae	0	1
Glypta fumiferanae ichnovirus	Ichnovirus	Polydnaviridae	0	2
Rotavirus sp.	Rotavirus	Reoviridae	0	1
Rotavirus A	Rotavirus	Reoviridae	0	67
Feline leukemia virus	Gammaretrovirus	Retroviridae	0	2
Human endogenous retrovirus		Retroviridae	0	3
Human endogenous retrovirus W		Retroviridae	0	3
Human endogenous retrovirus H		Retroviridae	0	8
Murine leukemia virus	Gammaretrovirus	Retroviridae	0	9
Human endogenous retrovirus K		Retroviridae	1	5
Aravan virus	Lyssavirus	Rhabdoviridae	0	1
Cocal virus	Vesiculovirus	Rhabdoviridae	0	1
Drosophila affinis sigmavirus	Sigmavirus	Rhabdoviridae	0	1
Drosophila immigrans sigmavirus	Sigmavirus	Rhabdoviridae	0	1
Drosophila obscura sigmavirus	Sigmavirus	Rhabdoviridae	0	1
Isfahan virus	Vesiculovirus	Rhabdoviridae	0	1
Porton's virus		Rhabdoviridae	0	1
Tupaia virus		Rhabdoviridae	0	1
Flanders virus		Rhabdoviridae	0	14
Chandipura virus	Vesiculovirus	Rhabdoviridae	0	19
Obodhiang virus	Ephemerovirus	Rhabdoviridae	0	2
Muscina stabulans sigmavirus	Sigmavirus	Rhabdoviridae	0	20
Siniperca chuatsi rhabdovirus		Rhabdoviridae	0	20
Wongabel virus		Rhabdoviridae	0	24

Drosophila tristis sigmavirus	Sigmavirus	Rhabdoviridae	0	3
Scophthalmus maximus rhabdovirus		Rhabdoviridae	0	3
Vesicular stomatitis Indiana virus	Vesiculovirus	Rhabdoviridae	0	3
Vesicular stomatitis virus	Vesiculovirus	Rhabdoviridae	0	32
Kamese virus		Rhabdoviridae	0	36
Starry flounder rhabdovirus		Rhabdoviridae	0	4
Bahia Grande virus		Rhabdoviridae	0	5
Muir Springs virus		Rhabdoviridae	0	8
Drosophila melanogaster sigmavirus	Sigmavirus	Rhabdoviridae	1	0
Coastal Plains virus	Tibrovirus	Rhabdoviridae	1	12
Bovine ephemeral fever virus	Ephemerovirus	Rhabdoviridae	1	13
Adelaide River virus	Ephemerovirus	Rhabdoviridae	1	16
Ngaingan virus		Rhabdoviridae	1	22
Kotonkon virus		Rhabdoviridae	1	8
Tibrogargan virus	Tibrovirus	Rhabdoviridae	2	77

Supplemental Table S21. Viral output of translated nucleotide alignment to viral protein DB for a hemorrhagic fever NGS dataset (H).

Species	Genus	Family	Contigs #7	Reads #7
Oryctes rhinoceros virus			0	1
Nyamanini virus			0	14
Heliothis zea virus 1			0	3
Midway virus			0	3
Varanid adenovirus Leipzig 12 2007	Atadenovirus	Adenoviridae	0	1
Cyprinid herpesvirus 3	Cyprinivirus	Alloherpesviridae	0	1
TTV-like mini virus		Anelloviridae	0	10
SEN virus		Anelloviridae	0	2
Torque teno midi virus		Anelloviridae	0	6
Trichoplusia ni ascovirus 2c	Ascovirus	Ascoviridae	0	1
Diadromus pulchellus ascovirus 4a	Ascovirus	Ascoviridae	0	11
Cydia pomonella granulovirus	Betabaculovirus	Baculoviridae	0	1
Cryptophlebia leucotreta granulovirus	Betabaculovirus	Baculoviridae	0	2
Pseudaletia unipuncta granulovirus	Betabaculovirus	Baculoviridae	0	2
Orgyia leucostigma NPV	Alphabaculovirus	Baculoviridae	0	4
Autographa californica multiple nucleopolyhedrovirus	Alphabaculovirus	Baculoviridae	0	9
Banana streak virus	Badnavirus	Caulimoviridae	0	1
Macaca fascicularis	Macaca	Cercopithecidae	0	1
Anthurium mosaic-associated virus	Chrysovirus	Chrysoviridae	0	5
Alphacoronavirus 1	Alphacoronavirus	Coronaviridae	0	1
Zaire ebolavirus	Ebolavirus	Filoviridae	0	13
Dengue virus	Flavivirus	Flaviviridae	0	1
Hepatitis C virus	Hepacivirus	Flaviviridae	0	1
Bovine viral diarrhea virus 1	Pestivirus	Flaviviridae	0	5
Wheat dwarf virus	Mastrevirus	Geminiviridae	0	2
Equid herpesvirus 2	Percavirus	Herpesviridae	0	1
Human herpesvirus 1	Simplexvirus	Herpesviridae	0	1
Murid herpesvirus 1	Muromegalovirus	Herpesviridae	0	1
Bovine herpesvirus 4	Rhadinovirus	Herpesviridae	0	2
Equid herpesvirus 1	Varicellovirus	Herpesviridae	0	2
Human herpesvirus 3	Varicellovirus	Herpesviridae	0	2
Human herpesvirus 8	Rhadinovirus	Herpesviridae	0	2
Human herpesvirus 6	Roseolovirus	Herpesviridae	0	27
Tupaia herpesvirus 1		Herpesviridae	0	3
Glossina hytivirus	Glossinavirus	Hytrosaviridae	0	1
Musca hytivirus	Muscavirus	Hytrosaviridae	0	1
Invertebrate iridescent virus 6	Iridovirus	Iridoviridae	0	4
Invertebrate iridescent virus 3	Chloriridovirus	Iridoviridae	0	5
Lymphocystis disease virus - isolate China	Lymphocystivirus	Iridoviridae	0	62
Acanthamoeba polyphaga mimivirus	Mimivirus	Mimiviridae	0	19
Moumouvirus		Mimiviridae	0	3
Courdo11 virus		Mimiviridae	0	8
White spot syndrome virus	Whispovirus	Nimaviridae	0	2
Alphapapillomavirus 10	Alphapapillomavirus	Papillomaviridae	0	2
Alphapapillomavirus 7	Alphapapillomavirus	Papillomaviridae	0	2
Feldmannia species virus	Phaeovirus	Phycodnaviridae	0	1
Ostreococcus tauri virus 1		Phycodnaviridae	0	1
Paramecium bursaria Chlorella virus NY2A	Chlorovirus	Phycodnaviridae	0	1
Acanthocystis turfacea Chlorella virus 1	Chlorovirus	Phycodnaviridae	0	12
Feldmannia irregularis virus a	Phaeovirus	Phycodnaviridae	0	13
Ectocarpus siliculosus virus 1	Phaeovirus	Phycodnaviridae	0	19
Ostreococcus virus OsV5		Phycodnaviridae	0	2
Paramecium bursaria Chlorella virus 1	Chlorovirus	Phycodnaviridae	0	2

Bathycoccus sp. RCC1105 virus BpV	Prasinovirus	Phycodnaviridae	0	27
Emiliana huxleyi virus 86	Coccolithovirus	Phycodnaviridae	0	3
Paramecium bursaria Chlorella virus A1	Chlorovirus	Phycodnaviridae	0	3
Emiliana huxleyi virus 99B1	Coccolithovirus	Phycodnaviridae	0	4
Ostreococcus lucimarinus virus OIV1	Prasinovirus	Phycodnaviridae	0	5
Paramecium bursaria Chlorella virus AR158	Chlorovirus	Phycodnaviridae	0	6
Micromonas sp. RCC1109 virus MpV1	Prasinovirus	Phycodnaviridae	0	7
Human picobirnavirus	Picobirnavirus	Picobirnaviridae	0	5
Encephalomyocarditis virus	Cardiovirus	Picornaviridae	0	2
Enterovirus B	Enterovirus	Picornaviridae	0	7
Glypta fumiferanae ichnovirus	Ichnovirus	Polydnaviridae	0	54
Abelson murine leukemia virus	Gammaretrovirus	Retroviridae	0	1
AKT8 retrovirus		Retroviridae	0	1
Avian carcinoma virus	Alpharetrovirus	Retroviridae	0	1
Mason-Pfizer monkey virus	Betaretrovirus	Retroviridae	0	1
Mouse mammary tumor virus	Betaretrovirus	Retroviridae	0	1
Rabbit endogenous retrovirus H	Betaretrovirus	Retroviridae	0	1
Simian retrovirus 4	Betaretrovirus	Retroviridae	0	1
Woolly monkey sarcoma virus	Gammaretrovirus	Retroviridae	0	1
Feline leukemia virus	Gammaretrovirus	Retroviridae	0	11
Atlantic salmon swim bladder sarcoma virus		Retroviridae	0	2
Enzootic nasal tumour virus of goats	Betaretrovirus	Retroviridae	0	2
RD114 retrovirus	Gammaretrovirus	Retroviridae	0	2
Reticuloendotheliosis virus	Gammaretrovirus	Retroviridae	0	2
Walleye dermal sarcoma virus	Epsilonretrovirus	Retroviridae	0	2
Porcine type-C oncovirus	Gammaretrovirus	Retroviridae	0	21
Human endogenous retrovirus H		Retroviridae	0	217
Murine leukemia-related retroviruses	Gammaretrovirus	Retroviridae	0	3
Baboon endogenous virus	Gammaretrovirus	Retroviridae	0	4
Human immunodeficiency virus 1	Lentivirus	Retroviridae	0	4
Human endogenous retrovirus K		Retroviridae	0	47
Murine leukemia virus	Gammaretrovirus	Retroviridae	0	5
Gibbon ape leukemia virus	Gammaretrovirus	Retroviridae	0	6
Human endogenous retrovirus W		Retroviridae	0	83
Human endogenous retrovirus		Retroviridae	0	93
Anguillid rhabdovirus	Perhabdovirus	Rhabdoviridae	0	1
Kimberley virus	Ephemerovirus	Rhabdoviridae	0	1
Kotonkan virus	Ephemerovirus	Rhabdoviridae	0	1
Oak-Vale virus		Rhabdoviridae	0	1
Drosophila tristis sigmavirus	Sigmavirus	Rhabdoviridae	0	10
Porton's virus		Rhabdoviridae	0	10
Scophthalmus maximus rhabdovirus		Rhabdoviridae	0	118
Vesicular stomatitis Alagoas virus	Vesiculovirus	Rhabdoviridae	0	12
Bovine ephemeral fever virus	Ephemerovirus	Rhabdoviridae	0	150
Pike fry rhabdovirus	Vesiculovirus	Rhabdoviridae	0	17
Vesicular stomatitis virus	Vesiculovirus	Rhabdoviridae	0	172
Australian bat lyssavirus	Lyssavirus	Rhabdoviridae	0	2
Drosophila ananassae sigmavirus	Sigmavirus	Rhabdoviridae	0	2
Drosophila immigrans sigmavirus	Sigmavirus	Rhabdoviridae	0	2
Duvenhage virus	Lyssavirus	Rhabdoviridae	0	2
Kern Canyon virus		Rhabdoviridae	0	2
Mossuril virus		Rhabdoviridae	0	2
Moussa virus		Rhabdoviridae	0	20
Obodhiang virus	Ephemerovirus	Rhabdoviridae	0	21
Adelaide River virus	Ephemerovirus	Rhabdoviridae	0	211
Tupaia virus		Rhabdoviridae	0	22
Ngaingan virus		Rhabdoviridae	0	277
Flanders virus		Rhabdoviridae	0	294
European bat lyssavirus 1	Lyssavirus	Rhabdoviridae	0	3
Spring viraemia of carp virus	Vesiculovirus	Rhabdoviridae	0	3
China fish rhabdovirus QZ-2005		Rhabdoviridae	0	30
Durham virus		Rhabdoviridae	0	30

Lagos bat virus	Lyssavirus	Rhabdoviridae	0	32
Coastal Plains virus	Tibrovirus	Rhabdoviridae	0	321
Bahia Grande virus		Rhabdoviridae	0	33
Berrimah virus	Ephemerovirus	Rhabdoviridae	0	33
Chandipura virus	Vesiculovirus	Rhabdoviridae	0	34
Mokola virus	Lyssavirus	Rhabdoviridae	0	35
Kamese virus		Rhabdoviridae	0	37
Muscina stabulans sigmavirus	Sigmavirus	Rhabdoviridae	0	417
Drosophila affinis sigmavirus	Sigmavirus	Rhabdoviridae	0	5
Isfahan virus	Vesiculovirus	Rhabdoviridae	0	5
Rabies virus	Lyssavirus	Rhabdoviridae	0	5
Aravan virus	Lyssavirus	Rhabdoviridae	0	51
Muir Springs virus		Rhabdoviridae	0	57
Siniperca chuatsi rhabdovirus		Rhabdoviridae	0	59
Kotonkon virus		Rhabdoviridae	0	68
Vesicular stomatitis Indiana virus	Vesiculovirus	Rhabdoviridae	0	69
Starry flounder rhabdovirus		Rhabdoviridae	0	78
Drosophila obscura sigmavirus	Sigmavirus	Rhabdoviridae	0	9
Cocal virus	Vesiculovirus	Rhabdoviridae	0	93
Drosophila melanogaster sigmavirus	Sigmavirus	Rhabdoviridae	0	93
Wongabel virus		Rhabdoviridae	1	214
Tibrogargan virus	Tibrovirus	Rhabdoviridae	2	928
Maize chlorotic dwarf virus	Waikavirus	Secoviridae	1	0
Phlebiopsis gigantea mycovirus dsRNA 1		Totiviridae	0	2

Supplemental Table S22. Performance testing of *de novo* assembly methods.

Abbreviations: k, kmer value; #, number of contigs generated that mapped to the targeted reference viral genome; %, coverage to targeted reference viral genome; N50, N50 statistic; AbySS-100k, *de novo* assembly running AbySS on individually partitioned sets of 100,000 reads.

Dataset	Targeted Virus	SOAP-denovo (k=35)		ABYSS (k=34)		AbySS-100k (k=34)		ABYSS 100k (k=34) + Minimo	
		% (#)	N50	% (#)	N50	% (#)	N50	% (#)	N50
B. pediatric diarrhea (stool)	SaV	52.5 (11)	306	11.2 (3)	274	96.7 (240)	275	98.3 (39)	408
B. pediatric diarrheal(stool)	HPeV-1	17.9 (7)	208	10.7 (5)	216	40.5 (23)	242	62.4 (25)	285
D. hemorrhagic fever (HiSeq,serum)	BASV	83 (20)	760	53.8 (8)	1525	86.1 (51)	281	86 (16)	895
H. hemorrhagic fever (MiSeq,serum)	BASV	8.2 (2)	594	0 (0)	—	100 (76)	736	100 (3)	9036
E. TTVS hepatitis (serum)	HCV-1b	0 (0)	—	0 (0)	—	76.5 (78)	240	87.7 (42)	319
G. encephalitis (CSF)	HHV-3	0 (0)	—	0 (0)	—	0 (0)	—	0 (0)	—
J. prostate cancer (cell line)	HPV-18	6.7 (2)	426	0 (0)	—	8.5 (3)	567	9.7 (3)	618
K. simian pneumonia (lung swab/tissue)	TMAv	53.6 (66)	363	33.8 (42)	339	56.3 (77)	326	56.7 (67)	383
L. spiked HIV (plasma) 10 ² copies/mL	HIV-1	0 (0)	—	0 (0)	—	0(0)	—	0 (0)	—
L. spiked HIV (plasma) 10 ³ copies/mL	HIV-1	0 (0)	—	6.7 (2)	324	6.8 (4)	325	8.2 (2)	457
L. spiked HIV (plasma) 10 ⁴ copies/mL	HIV-1	8.6 (2)	468	8.2 (2)	456	8.2 (3)	457	91.8 (8)	4369

SUPPLEMENTAL REFERENCES

- Abdeldaim GM, Stralin K, Olcen P, Blomberg J, Molling P, Herrmann B. 2013. Quantitative fuck gene polymerase chain reaction on sputum and nasopharyngeal secretions to detect Haemophilus influenzae pneumonia. *Diagnostic microbiology and infectious disease* **76**(2): 141-146.
- Akobeng AK. 2007. Understanding diagnostic tests 3: Receiver operating characteristic curves. *Acta Paediatr* **96**(5): 644-647.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* **215**(3): 403-410.
- Bhaduri A, Qu K, Lee CS, Ungewickell A, Khavari PA. 2012. Rapid identification of non-human sequences in high-throughput sequencing datasets. *Bioinformatics* **28**(8): 1174-1175.
- Castellarin M, Warren RL, Freeman JD, Dreolini L, Krzywinski M, Strauss J, Barnes R, Watson P, Allen-Vercoe E, Moore RA et al. 2012. Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. *Genome Res* **22**(2): 299-306.
- Chen EC, Yagi S, Kelly KR, Mendoza SP, Tarara RP, Canfield DR, Maninger N, Rosenthal A, Spinner A, Bales KL et al. 2011. Cross-species transmission of a novel adenovirus associated with a fulminant pneumonia outbreak in a new world monkey colony. *PLoS Pathog* **7**(7): e1002155.
- Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B et al. 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**(11): 1422-1423.
- de Hoon MJ, Imoto S, Nolan J, Miyano S. 2004. Open source clustering software. *Bioinformatics* **20**(9): 1453-1454.
- Grard G, Fair JN, Lee D, Slikas E, Steffen I, Muyembe JJ, Sittler T, Veeraraghavan N, Ruby JG, Wang C et al. 2012. A novel rhabdovirus associated with acute hemorrhagic fever in central Africa. *PLoS Pathog* **8**(9): e1002924.
- Gremme G, Steinbiss S, Kurtz S. 2013. GenomeTools: A Comprehensive Software Library for Efficient Processing of Structured Genome Annotations. *IEEE/ACM Trans Comput Biol Bioinform* **10**(3): 645-656.
- Greninger AL, Chen EC, Sittler T, Scheinerman A, Roubinian N, Yu G, Kim E, Pillai DR, Guyard C, Mazzulli T et al. 2010. A metagenomic analysis of pandemic influenza A (2009 H1N1) infection in patients from North America. *PLoS One* **5**(10): e13381.
- Hartline J, Mierek C, Knutson T, Kang C. 2013. Hantavirus infection in North America: a clinical review. *Am J Emerg Med* **31**(6): 978-982.
- Hsiang MS, Lin M, Dokomajilar C, Kemere J, Pilcher CD, Dorsey G, Greenhouse B. 2010. PCR-based pooling of dried blood spots for detection of malaria parasites: optimization and application to a cohort of Ugandan children. *Journal of clinical microbiology* **48**(10): 3539-3543.
- Kostic AD, Ojesina AI, Pedamallu CS, Jung J, Verhaak RG, Getz G, Meyerson M. 2011. PathSeq: software to identify or discover microbes by deep sequencing of human tissue. *Nat Biotechnol* **29**(5): 393-396.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**(4): 357-359.
- Lee D, Das Gupta J, Gaughan C, Steffen I, Tang N, Luk KC, Qiu X, Urisman A, Fischer N, Molinaro R et al. 2012. In-depth investigation of archival and prospectively collected samples reveals no evidence for XMRV infection in prostate cancer. *PLoS One* **7**(9): e44954.
- Li D, Schmaljohn AL, Anderson K, Schmaljohn CS. 1995. Complete nucleotide sequences of the M and S segments of two hantavirus isolates from California: evidence for

- reassortment in nature among viruses related to hantavirus pulmonary syndrome. *Virology* **206**(2): 973-983.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**(14): 1754-1760.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**(16): 2078-2079.
- Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, Li Y, Li S, Shan G, Kristiansen K et al. 2010. De novo assembly of human genomes with massively parallel short read sequencing. *Genome research* **20**(2): 265-272.
- Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal* **17**:1
- Michael NL, Herman SA, Kwok S, Dreyer K, Wang J, Christopherson C, Spadaro JP, Young KK, Polonis V, McCutchan FE et al. 1999. Development of calibrated viral load standards for group M subtypes of human immunodeficiency virus type 1 and performance of an improved AMPLICOR HIV-1 MONITOR test with isolates of diverse subtypes. *J Clin Microbiol* **37**(8): 2557-2563.
- Mosley JW, Aach RD, Hollinger FB, Stevens CE, Barbosa LH, Nemo GJ, Holland PV, Bancroft WH, Zimmerman HJ, Kuo G et al. 1990. Non-A, non-B hepatitis and antibody to hepatitis C virus. *JAMA* **263**(1): 77-78.
- Naccache SN, Greninger AL, Lee D, Coffey LL, Phan T, Rein-Weston A, Aronsohn A, Hackett J, Jr., Delwart EL, Chiu CY. 2013. The perils of pathogen discovery: origin of a novel parvovirus-like hybrid genome traced to nucleic Acid extraction spin columns. *J Virol* **87**(22): 11966-11977.
- Naeem R, Rashid M, Pain A. 2013. READSCAN: a fast and scalable pathogen discovery program with accurate genome relative abundance estimation. *Bioinformatics* **29**(3): 391-392.
- Nunez JJ, Fritz CL, Knust B, Buttke D, Enge B, Novak MG, Kramer V, Osadebe L, Messenger S, Albarino CG et al. 2014. Hantavirus Infections among Overnight Visitors to Yosemite National Park, California, USA, 2012. *Emerg Infect Dis* **20**(3): 386-393.
- Prensner JR, Iyer MK, Balbin OA, Dhanasekaran SM, Cao Q, Brenner JC, Laxman B, Asangani IA, Grasso CS, Kominsky HD et al. 2011. Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. *Nat Biotechnol* **29**(8): 742-749.
- Santos C, Reis A, Dos Santos CV, Damas C, Silva MH, Viana MV, Ferraz ML, Carnauba D, El-Far F, Serra F et al. 2012. The use of real-time PCR to detect hepatitis C virus RNA in dried blood spots from Brazilian patients infected chronically. *Journal of virological methods* **179**(1): 17-20.
- Schmaljohn AL, Li D, Negley DL, Bressler DS, Turell MJ, Korch GW, Ascher MS, Schmaljohn CS. 1995. Isolation and initial characterization of a newfound hantavirus from California. *Virology* **206**(2): 963-972.
- Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**(6): 863-864.
- Shendure J, Ji H. 2008. Next-generation DNA sequencing. *Nat Biotechnol* **26**(10): 1135-1145.
- Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I. 2009. ABySS: a parallel assembler for short read sequence data. *Genome Res* **19**(6): 1117-1123.
- Swei A, Russell BJ, Naccache SN, Kabre B, Veeraraghavan N, Pilgard MA, Johnson BJ, Chiu CY. 2013. The genome sequence of Lone Star virus, a highly divergent bunyavirus found in the Amblyomma americanum tick. *PLoS One* **8**(4): e62083.
- Treangen TJ, Sommer DD, Angly FE, Koren S, Pop M. 2011. Next generation sequence assembly with AMOS. *Curr Protoc Bioinformatics* **Chapter 11**: Unit 11 18.

- Wang D, Urisman A, Liu YT, Springer M, Ksiazek TG, Erdman DD, Mardis ER, Hickenbotham M, Magrini V, Eldred J et al. 2003. Viral discovery and sequence recovery using DNA microarrays. *PLoS biology* **1**(2): E2.
- Wilson MR, Naccache SN, Samayoa E, Biagtan M, Bashir H, Yu G, Salamat SM, Somasekar S, Federman S, Miller S et al. 2014. Actionable diagnosis of neuroleptospirosis by next-generation sequencing. *New England Journal of Medicine* **(in press)**.
- Yu G, Greninger AL, Isa P, Phan TG, Martinez MA, de la Luz Sanchez M, Contreras JF, Santos-Preciado JI, Parsonnet J, Miller S et al. 2012. Discovery of a novel polyomavirus in acute diarrheal samples from children. *PLoS One* **7**(11): e49449.
- Zaharia M, Bolosky B, Curtis K, Patterson D, Fox A, Patterson D, Shenker S, Stoica I, Sittler T. 2012. Alignment in a SNAP: Cancer Diagnosis in the Genomic Age. *Laboratory Investigation* **92**: 458A-458A.
- Zhao Y, Tang H, Ye Y. 2012. RAPSearch2: a fast and memory-efficient protein similarity search tool for next-generation sequencing data. *Bioinformatics* **28**(1): 125-126.