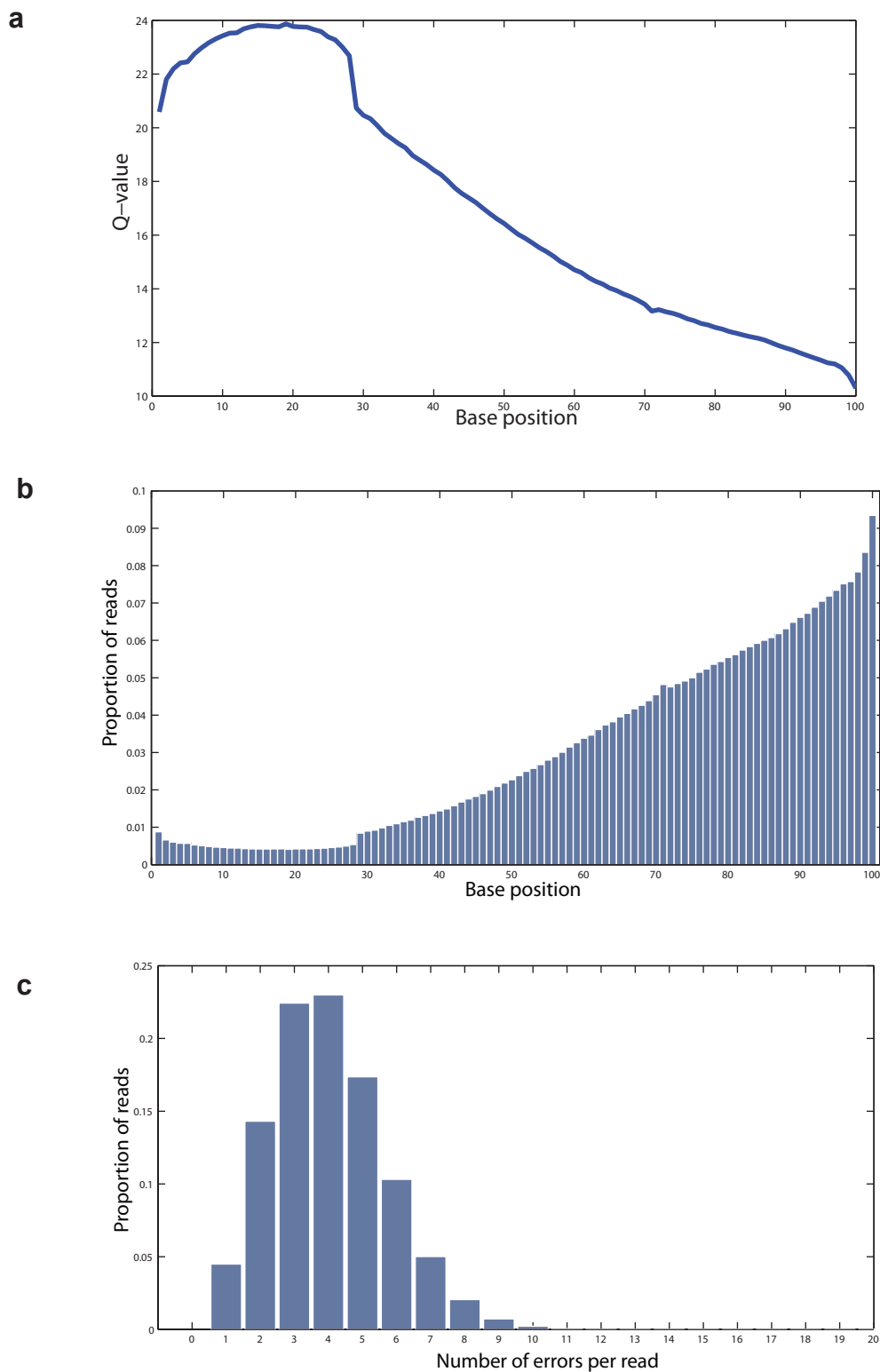
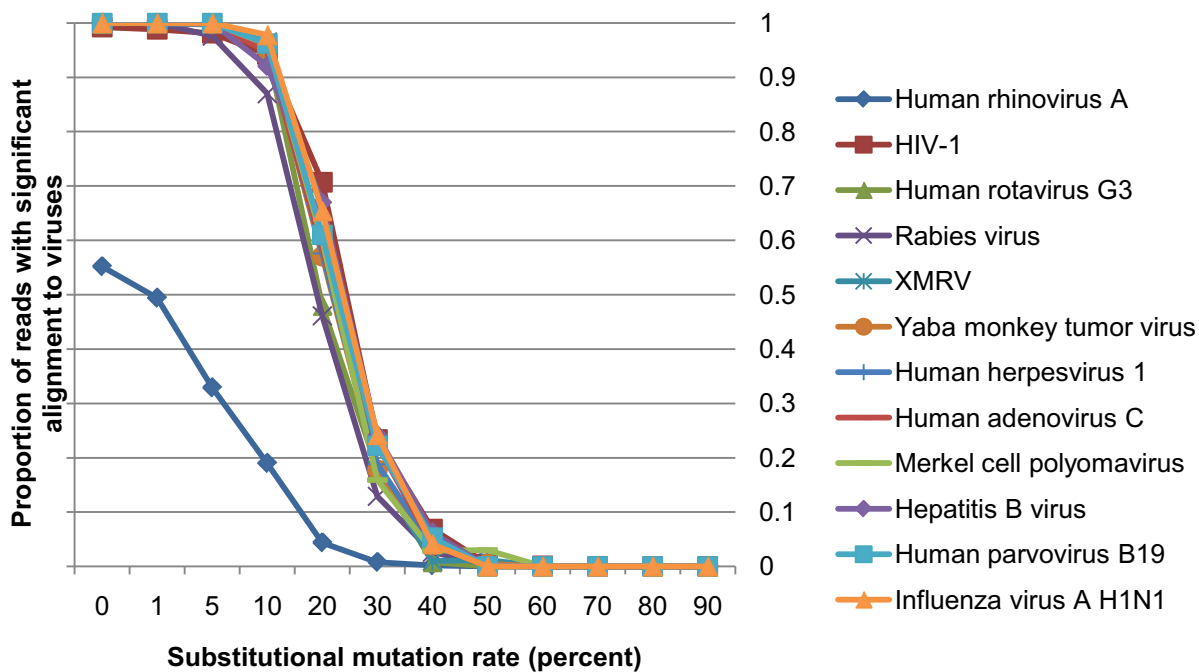
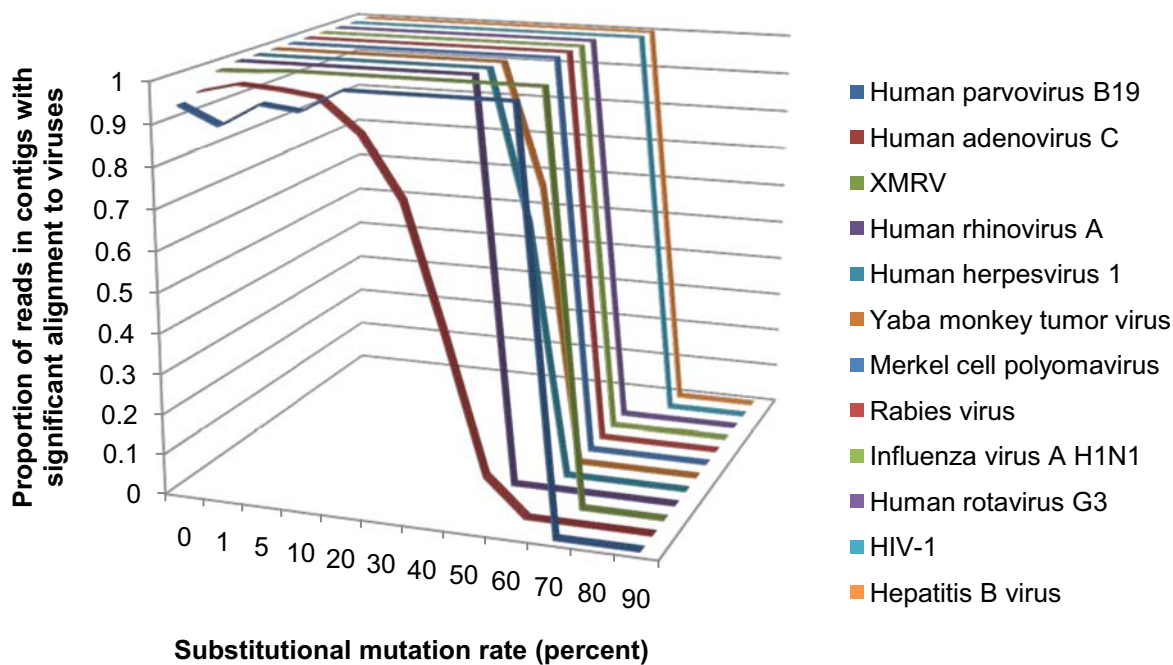


Supplementary Figure 1. Generation of artificial shotgun sequencing reads. (a) 20 million reads were generated from a human transcriptome database by randomly selecting 100-mer sequences. (b) A set of twelve virus genomes were selected (see **Methods**). For each genome, substitutional mutations were introduced at twelve distinct mutation rates, and for each of these mutated genomes 100-mer sequences were chosen at random to produce the read set.

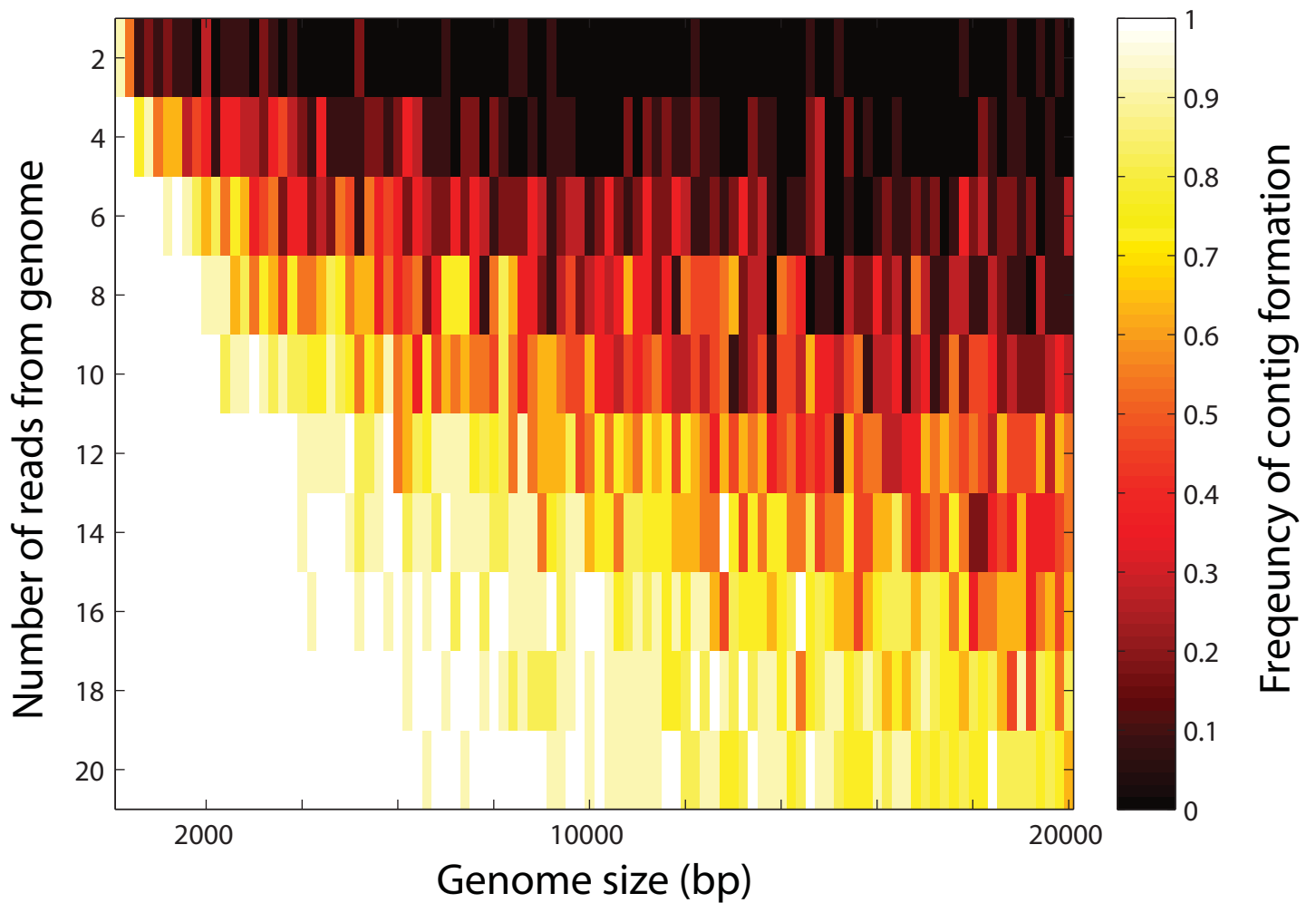


Supplementary Figure 2. Applying a sequencing error distribution model to artificially generated reads. (a) Average quality score plot for a typical set of reads generated by the Illumina GAII Sequencer. The Q-value is defined as $Q = -10 \cdot \log_{10}(p/(1-p))$, where p is the probability that the corresponding base call is incorrect. (b) The error distribution was applied to the complete set of reads generated as shown in **Supplementary Figure 1**. The plot shows the proportion of the read set containing a substituted base at the indicated position. (c) The proportion of the read set is shown as a function of the number of sequencing errors per read.

a**b**

Supplementary Figure 3. Identification of virus-derived reads by sequence similarity. (a) Reads were generated as indicated in **Supplementary Figure 1b**. The proportion of reads identified as a virus sequence (MegaBlast alignment, E-value < 10e-10) is shown as a function of the substitutional mutation rate of the mutated genome. (b) Independent assemblies were performed on all reads from each of the 144 genomes. The proportion of reads incorporated into contigs that were identified as a virus sequence (MegaBlast alignment, E-value < 10e-10) is shown.

Supplementary Figure 3



Supplementary Figure 4. Probability of contig formation for sequences randomly generated from a genome of varying size. This figure shows the probability of forming contigs from a set of simulated, virus-derived reads. Reads were generated by selecting random 100-mers from a series of genomes ranging in size from 100bp to 20,000bp, producing between 2 and 20 reads per genome. The heatmap indicates the frequency among 11 replicates with which a contig of at least 175bp was formed from these reads by *de novo* assembly.

Supplementary Table 1. Bacterial genomes used to construct an artificial sequence dataset to test the metagenomics module of PathSeq. Shown is the set of bacterial genomes that were each used to create a set of 10,000 random 100-mers. The species were chosen to represent both evolutionary relatedness and divergence to test the ability of the metagenomics module to properly assess the microbial representation of a mixed sample.

Accession	Definition
NC_003228.3	<i>Bacteroides fragilis</i> NCTC 9343, complete genome.
NC_009614.1	<i>Bacteroides vulgatus</i> ATCC 8482, complete genome.
NC_013316.1	<i>Clostridium difficile</i> R20291, complete genome.
NC_009615.1	<i>Parabacteroides distasonis</i> ATCC 8503, complete genome.
NC_004663.1	<i>Bacteroides thetaiotaomicron</i> VPI-5482, complete genome.
NC_010816.1	<i>Bifidobacterium longum</i> DJO10A, complete genome.
NC_008618.1	<i>Bifidobacterium adolescentis</i> ATCC 15703, complete genome.
NC_012781.1	<i>Eubacterium rectale</i> ATCC 33656, complete genome.
NC_012778.1	<i>Eubacterium eligens</i> ATCC 27750, complete genome.
NC_010655.1	<i>Akkermansia muciniphila</i> ATCC BAA-835, complete genome.
NC_011353.1	<i>Escherichia coli</i> O157:H7 str. EC4115, complete genome.
NC_004668.1	<i>Enterococcus faecalis</i> V583, complete genome.

Supplementary Table 2. Metagenomic analysis on a sequence dataset constructed from a set of twelve bacterial genomes. Shown is the number of reads that were identified as matching to the indicated bacterial genome (actual number of reads is 10,000 for each species).

Genus	Species	Number of Reads	Fraction Genome Coverage
Bifidobacterium	adolescentis	9877	0.376091154
Eubacterium	eligens	9903	0.368491132
Bifidobacterium	longum	9933	0.333497209
Akkermansia	muciniphila	9954	0.310035802
Enterococcus	faecalis	9900	0.264391797
Eubacterium	rectale	9953	0.249987753
Clostridium	difficile	9829	0.208571056
Parabacteroides	distasonis	9929	0.186090932
Bacteroides	vulgatus	9943	0.173829585
Bacteroides	fragilis	9929	0.169107843
Bacteroides	thetaitotaomicron	9933	0.146473822
Escherichia	coli	5133	0.089648584
Staphylococcus	aureus	3	1.04E-04
Bifidobacterium	dentium	1	3.79E-05
Shigella	dysenteriae	1	2.29E-05
Shigella	sonnei	1	2.07E-05

Supplementary Table 3. PathSeq performance on artificially generated sequence data. Reads were generated by sampling random 100-mer sequences from a human transcriptome database, generating 20 million reads, or from a set of twelve virus genomes each substitutionally mutated at twelve distinct rates, generating 144,000 reads. The rows represent the number of reads remaining at each step in the PathSeq workflow.

Stage in workflow	Human reads remaining	Virus reads remaining	Virus reads subtracted at each step
START	20000000	144000	
Duplicate Remover	19633552	144000	0
Maq database1	5441980	144000	0
Maq database2	696074	144000	0
Maq database3	1218	144000	0
Maq database4	1213	144000	0
RepeatMasker Remover	853	142880	1120
MegaBlast database1	0	142878	2
MegaBlast database2	0	142878	0
Blast database1	0	142878	0
Blast database2	0	142878	0

Supplementary Table 4. PathSeq performance on artificially generated sequence data with introduced sequencing errors. Reads are the same as in Table 1 except that “sequence errors” were introduced into the reads according to a sequencing error distribution model.

Stage in workflow	Human reads remaining	Virus reads remaining	Virus reads subtracted at each step
START	20000000	144000	
Duplicate Remover	19999188	144000	0
Maq database1	14240412	144000	0
Maq database2	12205718	144000	0
Maq database3	11745641	144000	0
Maq database4	11744303	144000	0
RepeatMasker Remover	11149237	143071	929
MegaBlast database1	54343	143070	1
MegaBlast database2	54329	143070	0
Blast database1	0	143069	1
Blast database2	0	143069	0

SUPPLEMENTARY METHODS

RNA-seq library construction from HeLa cells and Sequencing

RNA was extracted from cultured HeLa cells according to the RNeasy Kit (Qiagen) protocol. cDNA sequencing library construction was performed as described previously¹, with noted modifications below. The cDNA library was sequenced on the Illumina Genome Analyzer II (GAII) platform. The mean fragment length was approximately 350 base pairs. One lane of paired-end, 76 base pair sequencing was performed, producing 38.5 million purity filtered reads, which yielded 10,304,513 high quality reads following quality filtering.

Modifications made to the Illumina RNA-Seq Protocol

Total RNA (500 ng) was heated at 98°C for 100 min in THE RNA Storage Solution (1 mM sodium citrate, pH 6.4; Ambion/ABI, AM7000) to fragment the RNA to a mean size of ~500 nucleotides. Quality of RNA fragmentations was assessed on a Bioanalyzer 2100 (Agilent). First-strand cDNA synthesis was performed by adding random hexamers (Invitrogen, 48190-011) to the RNA and heating at 70°C for 10 min, and then immediately incubating at 50 °C for 1 h upon addition of Superscript III reverse transcriptase (Invitrogen). Second-strand synthesis was carried-out with *E. coli* DNA ligase and *E. coli* DNA polymerase I (Invitrogen) for 2.5 h at 16 °C. cDNA was purified using the MiniElute PCR Purification Kit (Qiagen) and evaluated using Bioanalyzer. End-repair, addition of adenine to the 3' end of the DNA fragments, and adapter ligation was performed as described in Guttman *et al.*, except a 2:1 molar ratio of adapter to DNA fragment was used during adapter-ligation. The resulting adapter-ligated fragments were

purified on a 4% SeaKem LE agarose gel (Lonza) and a 400-500 base pair band was cut out of the gel and purified using the MiniElute kit. PCR was performed with Phusion DNA polymerase (Finnzymes) and adapter-specific primers using the following conditions: 2 min at 98 °C; [10 s at 98 °C, 30 s at 65 °C, 30 s at 72 °C] for 13 cycles; 5 min at 72 °C. Following PCR, a second round of gel extraction was performed as described above, and the product was submitted for Illumina sequencing.

The PathSeq workflow and cloud implementation

The PathSeq pipeline is designed using the Apache Hadoop implementation of the MapReduce programming framework (<http://hadoop.apache.org/mapreduce>) and can be run on the Amazon Elastic Compute Cloud (EC2) (<http://aws.amazon.com/ec2/>)². The workflow is comprised of three modules: pre-subtraction, subtraction, and post-subtraction. The pre-subtraction module is simply a quality filtering step and is run on the user's local machine, whereas the subtraction and post-subtraction modules are executed on a Hadoop-based cluster (19 worker nodes and 1 master node) built using the Amazon Elastic Compute Cloud (Amazon EC2).

Amazon's Simple Storage Service (S3) file system (<http://aws.amazon.com/s3/>) is used to store the reference sequences and readset, and the config files and scripts are distributed across all nodes on the cluster using the Hadoop Distributed File System (HDFS). The reference sequences are continuously updated on the PathSeq system and users are given the option of substituting any built-in database with a database of their choice; however we provide data download dates for reference sequences used in experiments reported in this paper below.

All processes are run on the Hadoop cluster in multiple map phases. The subtraction module comprises of two mappers. First, subtractive alignments are performed with MAQ

(Release 0.5.0, default settings) against a set of six human sequence databases: the 1000 Genomes Project female reference (ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp/technical/reference/, downloaded 2009-04-11), the Ensembl *Homo sapiens* cDNA database (ftp://ftp.ensembl.org/pub/current/fasta/homo_sapiens/cdna/, downloaded 2009-04-22), the human genome and transcriptome BLAST database (ftp://ftp.ncbi.nih.gov/blast/db/, downloaded 2009-05-25), and the set of three assembled human genomes available on NCBI (hs_alt_Celera, hs_alt_HuRef, hs_ref_GRCh37, ftp://ftp.ncbi.nih.gov/genomes/H_sapiens/Assembled_chromosomes/, downloaded 2009-06-19). The next map phase is composed of three steps: RepeatMasker (<http://www.repeatmasker.org/>), MegaBlast, and BLASTN. First, the reads are applied to RepeatMasker (version open-3.2.8, libraries dated 2009-06-04), and any reads with three or more masked nucleotides are discarded. Subtractive alignments are next performed using MegaBlast (Blast Tools version 2.2.23, cut-off expect value 10^{-7} , word size 16) to two human sequence databases: the NCBI *Homo sapiens* RNA database (ftp://ftp.ncbi.nih.gov/genomes/H_sapiens/RNA/, downloaded 2009-11-20), and the Ensembl human genome reference (ftp://ftp.ensembl.org/pub/current/fasta/homo_sapiens/dna/, downloaded 2009-10-22). The final set of subtractive alignments are then performed with BLASTN (Blast Tools version 2.2.23, cut-off expect value 10^{-7} , word size 7, nucleotide match reward 1, nucleotide mismatch reward -3, gap open cost 5, gap extension cost 2) to the same two databases. A reduce phase gathers all remaining reads into one consolidated file which serves as input to the post-subtraction module.

The post-subtraction module is also comprised of two mappers. The first mapper is a set of BLASTN (parameters as above) and BLASTX alignments (Blast Tools version 2.2.23, cut-off expect value 10^{-4} , word size 3, matrix: BLOSUM62, gap open cost 11, gap extension cost 1) to

viral (downloaded from NCBI Nucleotide (<http://www.ncbi.nlm.nih.gov/nucleotide>) using search term “viruses'[porgn: __txid10239]” on 2010-02-26), fungal (downloaded using the term “fungi'[porgn: __txid4751]” on 2009-11-23), bacterial and archaeal (<ftp://ftp.ncbi.nih.gov/genomes/Bacteria/>, downloaded 2010-03-30), and non-redundant protein (<ftp://ftp.ncbi.nih.gov/blast/db/nr>, downloaded 2010-04-05) reference sequences. This mapper also performs a *de novo* assembly (Velvet 0.7.31, k-mer size 21) on the full set of reads remaining from the previous map phase. The complete post-subtraction BLAST output files as well as the full set of unmapped reads and contigs are then uploaded and stored on the S3 storage system.

The Amazon Machine Image (AMI) required to build the PathSeq Hadoop cluster is accessible from Amazon Web Services (<http://www.broadinstitute.org/software/pathseq/>). Pathseq is implemented in Python, Java, C++ and C shell, and has been tested on a Linux 2.6.18-194.8.1.e15 X86_64 system.

PathSeq runtime and performance

PathSeq analysis was performed using a cluster of 19 worker nodes and 1 master node, which were EC2 Large CPU instances (7GB of memory and 2 processor cores). Full analysis of HeLa cell RNA-Seq data described in this report was performed in approximately 13 hours (wall clock time) for a total price of \$89 USD. The CPU time for this analysis was approximately 270 hours. Actual runtime and cost may vary depending on congestion on the Amazon EC2, Internet traffic, and the method of data upload. Because of its parallel architecture, PathSeq can analyze substantially larger datasets in a similar timeframe simply by increasing the cluster size.

Metagenomic analysis

The metagenomic analysis module of PathSeq reports the relative abundance of bacteria and archaea. This analysis begins with a MegaBlast alignment of the readset against the complete set of fully sequenced bacterial and archaeal genomes (<ftp://ftp.ncbi.nih.gov/genomes/Bacteria/>, downloaded 2010-03-30), reporting all hits with >90% sequence identity and >90% query coverage. The top 30 hits for each read are reported. Using these alignment results, classifications of each read are attempted at the phylum, then genus, then species level. If a given read cannot be classified uniquely at a given classification level (i.e. it has multiple hits to different reference sequences with equivalent E-values), then it is considered ambiguous and discarded from analysis at that level. Using species-level classifications, the fraction-genome-coverage is calculated for each species that received a hit, and this metric is used to quantify the relative abundance of a given species, normalized by the genome size.

Generation and analysis of simulated sequencing data

Simulated human transcriptome and virus sequence data. Twenty million 100-mers were randomly generated from a reference human transcriptome (ftp://ftp.ncbi.nih.gov/genomes/H_sapiens/RNA/, downloaded 2009-11-20). Simulated virus reads were derived from twelve virus sequences: NCBI Nucleotide accession AY740741, CY000455, EU643590, FJ356716, FJ464337, GQ290462, GQ415051, NC_000883, NC_001405, NC_001806, NC_005179, and NC_007815. For each of these sequences, substitutional mutations were introduced at a frequency of 0%, 1%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, and 90%. In this process, nucleotides along the sequence are chosen at random with replacement (i.e. the same nucleotide can be chosen twice at random) and converted

to a different nucleotide (for example, A is converted to C, G, or T). This produces 144 sequences (12 input sequences, each mutated at 12 frequencies). For each of these 144 sequences, 1,000 “reads” of length 100bp were produced at random. The resulting 144,000 simulated reads were pooled with the 20 million simulated human reads and analyzed on PathSeq.

Contig formation simulations. In this experiment, “genomes” of size 200bp to 20,000bp in increments of 200bp were generated from the Human herpesvirus 5 genome (accession GQ466044). For each of these 100 “genomes”, two to twenty 100-mer sequences (“reads”) were generated at random. This process was performed eleven times. For each “genome” size, “read” number pair, Velvet assembly with k-mer size 21 was performed. The frequency with which contigs of at least 175bp were generated was recorded.

Introduction of “sequencing errors” into simulated sequence data

Sequencing errors were introduced into the simulated reads based on quality scores seen in a whole-genome sequence dataset of a glioblastoma multiforme (GBM) primary tumor (sequenced as part of The Cancer Genome Atlas; data available via the NCBI Short Read Archive (SRA) identifier SRX010704). The average quality score for each base along the length of the reads was calculated across the dataset and offset by -5 (**Supplementary Fig. 3a**). This was converted into a probability value and used to “mutate” our simulated reads (i.e. for a sequence error probability of 0.001, there is a 0.1% chance that the base will be converted to a different base).

Human whole-genome ovarian tumor sequencing data

The human ovarian tumor whole-genome sequencing dataset was sequenced as part of The Cancer Genome Atlas, and the data is available via the NCBI SRA identifier SRX010747. This is a 101 base pair, paired-end sequence dataset with a nominal fragment length of 264 base pairs.

REFERENCES

1. Guttman, M. et al. *Nat Biotechnol* **28**, 503-510 (2010).
2. Wall, D.P. et al. *BMC Bioinformatics* **11**, 259 (2010).

Supplementary Data Set 1. Microbial sequences identified among the unmapped clone end sequences from *Wheeler et al.* We performed a MegaBlast alignment of all contigs against bacterial, viral, and fungal nucleotide databases. The most significant alignments are reported.

Query name	Query length	BLAST Database	Hit ID	Bit score	E-value
contig freeze2_11177	914	Bacteria	gi 191636824 ref NC_010999.1	1431.17	0
contig freeze2_11275	999	Bacteria	gi 191636824 ref NC_010999.1	1594.6	0
contig freeze2_10679	1167	Bacteria	gi 191636824 ref NC_010999.1	2057.96	0
contig freeze2_30701	816	Bacteria	gi 198282148 ref NC_011206.1	712.082	0
contig freeze2_31723	737	Bacteria	gi 198282148 ref NC_011206.1	533.272	8.09E-149
contig freeze2_14303	866	Bacteria	gi 91774356 ref NC_007947.1	838.979	0
contig freeze2_10972	976	Bacteria	gi 116493574 ref NC_008526.1	1636.89	0
contig freeze2_11051	1117	Bacteria	gi 116493574 ref NC_008526.1	1792.63	0
contig freeze2_30191	850	Bacteria	gi 198282148 ref NC_011206.1	658.247	0
contig freeze2_10883	1429	Bacteria	gi 191636824 ref NC_010999.1	2400.2	0
contig freeze2_10866	1012	Bacteria	gi 116493574 ref NC_008526.1	1461.93	0
contig freeze2_10893	1018	Bacteria	gi 191636824 ref NC_010999.1	1236.98	0
contig freeze2_10847	1718	Bacteria	gi 191636824 ref NC_010999.1	2798.2	0
contig freeze2_29354	953	Bacteria	gi 198282148 ref NC_011206.1	633.252	8.43E-179
contig freeze2_10443	1376	Bacteria	gi 191636824 ref NC_010999.1	2223.31	0
contig freeze2_10302	1610	Bacteria	gi 116493574 ref NC_008526.1	2342.52	0
contig freeze2_10881	859	Bacteria	gi 258506995 ref NC_013198.1	1286.97	0
contig freeze2_10536	1560	Bacteria	gi 191636824 ref NC_010999.1	2384.82	0
contig freeze2_10792	1135	Bacteria	gi 191636824 ref NC_010999.1	1658.04	0
contig freeze2_10806	1223	Bacteria	gi 191636824 ref NC_010999.1	1909.92	0
contig freeze2_11056	1045	Bacteria	gi 116493574 ref NC_008526.1	1542.68	0
contig freeze2_10841	1408	Bacteria	gi 191636824 ref NC_010999.1	1992.59	0
contig freeze2_18067	798	Bacteria	gi 190572091 ref NC_010943.1	783.221	0
contig freeze2_10478	938	Bacteria	gi 191636824 ref NC_010999.1	863.974	0
contig freeze2_12874	596	Fungi	gi 162949218 ref NC_001139.8	346.771	4.33E-93
contig freeze2_12205	809	Fungi	gi 162949218 ref NC_001139.8	419.833	6.02E-115
contig freeze2_12696	727	Fungi	gi 162949218 ref NC_001139.8	398.684	1.26E-108

Supplementary Data Set 2. Microbial sequences identified among the novel sequences from the Asian and African genomes from *Li et al.* The set of novel sequences from the Asian and African genome were downloaded (<http://yh.genomics.org.cn/download.jsp#pd>) and aligned by MegaBlast against bacterial, viral and fungal nucleotide databases. The most significant alignments are reported.

Query name	Query length	BLAST Database	Hit ID	Bit score	E-value
C119303880_1_334.0:1-332	332	Viruses	gi 82503188 ref NC_007605.1	639.02	0
C119524968_1_364.0:1-362	362	Viruses	gi 82503188 ref NC_007605.1	696.7	0
C119600616_1_375.0:1-373	373	Viruses	gi 82503188 ref NC_007605.1	717.85	0
C119674423_1_386.0:1-384	384	Viruses	gi 82503188 ref NC_007605.1	739	0
C119705567_1_391.0:1-389	389	Viruses	gi 139424470 ref NC_009334.1	748.613	0
C119726695_1_395.0:1-393	393	Viruses	gi 139424470 ref NC_009334.1	756.304	0
C119731793_1_395.0:1-393	393	Viruses	gi 82503188 ref NC_007605.1	756.304	0
C119744189_1_397.0:1-395	395	Viruses	gi 82503188 ref NC_007605.1	760.149	0
C119748653_1_398.0:1-396	396	Viruses	gi 139424470 ref NC_009334.1	762.072	0
C119896026_1_423.0:1-421	421	Viruses	gi 82503188 ref NC_007605.1	806.294	0
C119902298_1_424.0:41-422	382	Viruses	gi 82503188 ref NC_007605.1	735.154	0
C119921199_1_427.0:1-425	425	Viruses	gi 82503188 ref NC_007605.1	817.83	0
C119940413_1_431.0:1-429	429	Viruses	gi 139424470 ref NC_009334.1	825.521	0
C120002643_1_442.0:1-440	440	Viruses	gi 139424470 ref NC_009334.1	846.67	0
C120149024_1_469.0:1-467	467	Viruses	gi 82503188 ref NC_007605.1	898.583	0
C120191480_1_478.0:1-476	476	Viruses	gi 139424470 ref NC_009334.1	915.887	0
C120258135_1_491.0:1-489	489	Viruses	gi 82503188 ref NC_007605.1	940.882	0
C120330447_1_507.0:1-505	505	Viruses	gi 82503188 ref NC_007605.1	971.645	0
C120346593_1_510.0:1-508	508	Viruses	gi 82503188 ref NC_007605.1	977.413	0
C120492119_1_544.0:1-542	542	Viruses	gi 139424470 ref NC_009334.1	719.773	0
C120522603_1_551.0:1-549	549	Viruses	gi 139424470 ref NC_009334.1	1056.24	0
C120531769_1_554.0:1-552	552	Viruses	gi 139424470 ref NC_009334.1	1040.86	0
C120703811_1_598.0:1-596	596	Viruses	gi 82503188 ref NC_007605.1	1135.07	0
C120935232_1_665.0:1-663	663	Viruses	gi 82503188 ref NC_007605.1	1275.43	0
C120951592_1_671.0:1-669	669	Viruses	gi 139424470 ref NC_009334.1	1254.28	0
C120960358_1_673.0:1-671	671	Viruses	gi 82503188 ref NC_007605.1	1290.81	0
C121286040_1_791.0:1-789	789	Viruses	gi 82503188 ref NC_007605.1	1517.69	0
C121288274_1_792.0:1-790	790	Viruses	gi 82503188 ref NC_007605.1	1506.15	0
C121293122_1_794.0:1-792	792	Viruses	gi 82503188 ref NC_007605.1	1517.69	0
C121319800_1_805.0:1-803	803	Viruses	gi 82503188 ref NC_007605.1	1544.61	0
C121320126_1_805.0:1-803	803	Viruses	gi 82503188 ref NC_007605.1	1544.61	0
C121347810_1_817.0:1-815	815	Viruses	gi 139424470 ref NC_009334.1	1554.22	0
C121428246_1_852.0:1-850	850	Viruses	gi 82503188 ref NC_007605.1	1634.97	0
C121453377_1_864.0:1-862	862	Viruses	gi 82503188 ref NC_007605.1	1171.6	0
C121497526_1_885.0:1-883	883	Viruses	gi 82503188 ref NC_007605.1	1698.42	0
C121554558_1_914.0:1-912	912	Viruses	gi 82503188 ref NC_007605.1	1738.8	0
C121571538_1_923.0:1-921	921	Viruses	gi 82503188 ref NC_007605.1	1769.56	0
C121651310_1_966.0:1-964	964	Viruses	gi 82503188 ref NC_007605.1	1854.16	0
C121664332_1_973.0:1-971	971	Viruses	gi 82503188 ref NC_007605.1	1867.62	0
C121768448_1_1036.0:1-1034	1034	Viruses	gi 82503188 ref NC_007605.1	1988.75	0

C121866248_1_1102.0:1-1100	1100	Viruses	gi 82503188 ref NC_007605.1	2115.64	0
C121868598_1_1104.0:1-1102	1102	Viruses	gi 82503188 ref NC_007605.1	2119.49	0
C121889809_1_1119.0:1-1117	1117	Viruses	gi 82503188 ref NC_007605.1	2148.33	0
C121895511_1_1123.0:1-1121	1121	Viruses	gi 82503188 ref NC_007605.1	2156.02	0
C121913385_1_1137.0:1-1135	1135	Viruses	gi 82503188 ref NC_007605.1	2182.94	0
C121922687_1_1144.0:1-1142	1142	Viruses	gi 82503188 ref NC_007605.1	2192.55	0
C121941707_1_1158.0:1-1156	1156	Viruses	gi 82503188 ref NC_007605.1	2207.93	0
C122032117_1_1233.0:1-1231	1231	Viruses	gi 139424470 ref NC_009334.1	2352.13	0
C122181441_1_1381.0:1-1379	1379	Viruses	gi 139424470 ref NC_009334.1	2652.07	0
C122299164_1_1526.0:1-1524	1524	Viruses	gi 82503188 ref NC_007605.1	2930.86	0
C122328330_1_1567.0:1-1565	1565	Viruses	gi 82503188 ref NC_007605.1	3009.69	0
C122330180_1_1570.0:1-1568	1568	Viruses	gi 82503188 ref NC_007605.1	3015.46	0
C122362574_1_1620.0:1-1618	1618	Viruses	gi 82503188 ref NC_007605.1	3111.6	0
C122435276_1_1745.0:1-1743	1743	Viruses	gi 82503188 ref NC_007605.1	3351.93	0
C122441380_1_1757.0:1-1755	1755	Viruses	gi 82503188 ref NC_007605.1	3375	0
C122442982_1_1760.0:1-1758	1758	Viruses	gi 82503188 ref NC_007605.1	3380.77	0
C122455201_1_1784.0:1-1782	1782	Viruses	gi 82503188 ref NC_007605.1	3178.89	0
C122458643_1_1791.0:1-1789	1789	Viruses	gi 82503188 ref NC_007605.1	3440.38	0
C122460615_1_1795.0:1-1793	1793	Viruses	gi 82503188 ref NC_007605.1	3448.07	0
C122480117_1_1836.0:1-1834	1834	Viruses	gi 82503188 ref NC_007605.1	3526.9	0
C122499557_1_1880.0:1-1878	1878	Viruses	gi 82503188 ref NC_007605.1	3611.5	0
C122500603_1_1882.0:1-1880	1880	Viruses	gi 82503188 ref NC_007605.1	3615.34	0
C122532157_1_1960.0:1-1958	1958	Viruses	gi 82503188 ref NC_007605.1	3765.31	0
C122536387_1_1970.0:1-1968	1968	Viruses	gi 82503188 ref NC_007605.1	3782.61	0
C122640872_1_2310.0:1-2308	2308	Viruses	gi 82503188 ref NC_007605.1	4438.25	0
C122682056_1_2501.0:1-2499	2499	Viruses	gi 82503188 ref NC_007605.1	4805.48	0
C122683436_1_2508.0:1-2506	2506	Viruses	gi 82503188 ref NC_007605.1	4818.94	0
C122691220_1_2551.0:1-2549	2549	Viruses	gi 82503188 ref NC_007605.1	4892.01	0
C122698812_1_2596.0:1-2594	2594	Viruses	gi 82503188 ref NC_007605.1	4988.14	0
C122715178_1_2699.0:1-2697	2697	Viruses	gi 82503188 ref NC_007605.1	5186.18	0
C122720738_1_2737.0:1-2735	2735	Viruses	gi 82503188 ref NC_007605.1	5259.24	0
C122742268_1_2909.0:1-2907	2907	Viruses	gi 82503188 ref NC_007605.1	5589.94	0
C122747872_1_2961.0:1-2959	2959	Viruses	gi 82503188 ref NC_007605.1	5689.92	0
C122766940_1_3166.0:1-3164	3164	Viruses	gi 82503188 ref NC_007605.1	6078.3	0
C122770662_1_3216.0:1-3214	3214	Viruses	gi 82503188 ref NC_007605.1	6180.21	0
C122779453_1_3342.0:1-3340	3340	Viruses	gi 82503188 ref NC_007605.1	6422.47	0
C122807074_1_3934.0:1-3932	3932	Viruses	gi 82503188 ref NC_007605.1	7560.7	0
C122816684_1_4307.0:1-4305	4305	Viruses	gi 82503188 ref NC_007605.1	8264.4	0
C122823252_1_4697.0:1-4695	4695	Viruses	gi 82503188 ref NC_007605.1	9027.71	0
C122832902_1_6215.0:1-6213	6213	Viruses	gi 82503188 ref NC_007605.1	11946.3	0
C122833702_1_6620.0:1-6618	6618	Viruses	gi 82503188 ref NC_007605.1	12725	0
C122833934_1_6767.0:1-6765	6765	Viruses	gi 82503188 ref NC_007605.1	13007.7	0
scaffold168760_1_552.0:1-468	468	Viruses	gi 82503188 ref NC_007605.1	888.969	0
scaffold25509_1_5582.10:1-717	717	Viruses	gi 9626372 ref NC_001422.1	1379.25	0
scaffold25509_1_5582.12:1-874	874	Viruses	gi 9626372 ref NC_001422.1	1617.67	0
scaffold25509_1_5582.15:1-457	457	Viruses	gi 9626372 ref NC_001422.1	879.356	0
scaffold25509_1_5582.2:1-837	837	Viruses	gi 9626372 ref NC_001422.1	1609.98	0
C154349094_1_379.0:1-377	377	Viruses	gi 9626158 ref NC_001405.1	714.005	0

scaffold25509_1_5582.11:1-297	297	Viruses	gi 9626372 ref NC_001422.1	571.726	1.58E-162
C119074009_1_306.0:1-304	304	Viruses	gi 82503188 ref NC_007605.1	569.803	6.16E-162
C118836010_1_281.0:1-279	279	Viruses	gi 82503188 ref NC_007605.1	537.117	3.88E-152
C118994787_1_297.0:1-295	295	Viruses	gi 82503188 ref NC_007605.1	537.117	4.12E-152
C118747645_1_272.0:1-270	270	Viruses	gi 139424470 ref NC_009334.1	519.813	6.06E-147
C118700636_1_268.0:1-266	266	Viruses	gi 82503188 ref NC_007605.1	512.122	1.23E-144
scaffold25509_1_5582.6:1-264	264	Viruses	gi 9626372 ref NC_001422.1	490.973	2.84E-138
scaffold25509_1_5582.4:1-208	208	Viruses	gi 9626372 ref NC_001422.1	400.606	3.49E-111
C117842211_1_203.0:1-201	201	Viruses	gi 82503188 ref NC_007605.1	387.148	3.79E-107
C117849537_1_203.0:1-201	201	Viruses	gi 139424470 ref NC_009334.1	387.148	3.79E-107
scaffold25509_1_5582.7:1-197	197	Viruses	gi 9626372 ref NC_001422.1	379.457	7.65E-105
C117426534_1_182.0:1-180	180	Viruses	gi 139424470 ref NC_009334.1	346.771	4.78E-95
scaffold25509_1_5582.16:1-173	173	Viruses	gi 9626372 ref NC_001422.1	333.312	5.14E-91
scaffold25509_1_5582.5:1-150	150	Viruses	gi 9626372 ref NC_001422.1	289.091	8.97E-78
scaffold25509_1_5582.9:1-142	142	Viruses	gi 9626372 ref NC_001422.1	273.709	3.62E-73
C116026170_1_133.0:1-131	131	Viruses	gi 82503188 ref NC_007605.1	252.559	7.67E-67
C115756264_1_127.0:1-125	125	Viruses	gi 82503188 ref NC_007605.1	241.023	2.16E-63
C115761692_1_127.0:1-125	125	Viruses	gi 82503188 ref NC_007605.1	241.023	2.16E-63
scaffold62567_1_922.3:1-125	125	Viruses	gi 82503188 ref NC_007605.1	241.023	2.16E-63
scaffold25509_1_5582.14:1-124	124	Viruses	gi 9626372 ref NC_001422.1	239.101	8.10E-63
C151024180_1_126.0:1-124	124	Viruses	gi 9626243 ref NC_001416.1	235.255	1.16E-61
scaffold62567_1_922.4:1-119	119	Viruses	gi 82503188 ref NC_007605.1	229.487	6.04E-60
C154594504_1_424.0:1-422	422	Bacteria	gi 254667448 ref NC_013010.1	812.062	0
C154645116_1_434.0:1-432	432	Bacteria	gi 209916806 ref NC_011407.1	642.865	0
C155261927_1_583.0:1-581	581	Bacteria	gi 209921952 ref NC_011419.1	1112	0
C155828458_1_777.0:1-775	775	Bacteria	gi 152973607 ref NC_009650.1	1479.23	0
C153974801_1_321.0:1-319	319	Bacteria	gi 253750923 ref NC_012924.1	515.968	5.42E-144
C153399044_1_253.0:1-251	251	Bacteria	gi 253750923 ref NC_012924.1	471.746	8.56E-131
C152979378_1_217.0:1-215	215	Bacteria	gi 254667448 ref NC_013010.1	396.761	2.71E-108
C152733830_1_200.0:1-198	198	Bacteria	gi 254667448 ref NC_013010.1	381.38	1.05E-103
C152632972_1_193.0:1-191	191	Bacteria	gi 157149504 ref NC_009790.1	367.921	1.14E-99
C152435528_1_182.0:1-180	180	Bacteria	gi 254667448 ref NC_013010.1	346.771	2.47E-93
C152451524_1_183.0:1-181	181	Bacteria	gi 254667448 ref NC_013010.1	337.158	1.95E-90
C152085975_1_164.0:1-162	162	Bacteria	gi 18466424 ref NC_003384.1	312.163	5.74E-83
C151524210_1_142.0:1-140	140	Bacteria	gi 18466424 ref NC_003384.1	269.864	2.61E-70
C151497500_1_141.0:1-139	139	Bacteria	gi 152973607 ref NC_009650.1	267.941	9.82E-70
C151911791_1_156.0:1-154	154	Bacteria	gi 146317663 ref NC_009442.1	262.173	6.05E-68
C151376510_1_137.0:1-135	135	Bacteria	gi 254667448 ref NC_013010.1	260.25	1.96E-67
C151136665_1_130.0:1-128	128	Bacteria	gi 254667448 ref NC_013010.1	246.791	2.07E-63
C151162069_1_130.0:1-128	128	Bacteria	gi 254667448 ref NC_013010.1	241.023	1.13E-61
C151164119_1_130.0:1-128	128	Bacteria	gi 254667448 ref NC_013010.1	235.255	6.14E-60
C150834658_1_122.0:1-120	120	Bacteria	gi 254667448 ref NC_013010.1	231.41	8.16E-59
C150539079_1_115.0:1-113	113	Bacteria	gi 253750923 ref NC_012924.1	217.951	8.52E-55
C150516123_1_114.0:1-112	112	Bacteria	gi 219682499 ref NC_011835.1	216.028	3.19E-54
C150634133_1_117.0:1-115	115	Bacteria	gi 254667448 ref NC_013010.1	216.028	3.30E-54
C150437527_1_113.0:1-111	111	Bacteria	gi 254667448 ref NC_013010.1	214.106	1.20E-53
C150438245_1_113.0:1-111	111	Bacteria	gi 254667448 ref NC_013010.1	214.106	1.20E-53
C150340425_1_111.0:1-109	109	Bacteria	gi 254667448 ref NC_013010.1	210.26	1.68E-52

C149877309_1_103.0:1-101	101	Bacteria	gi 253750923 ref NC_012924.1	189.111	3.55E-46
C152535080_1_187.0:1-185	185	Bacteria	gi 146317663 ref NC_009442.1	285.245	8.48E-75
C151256372_1_133.0:1-131	131	Bacteria	gi 254667448 ref NC_013010.1	244.869	8.07E-63
C150829637_1_121.0:1-119	119	Bacteria	gi 157149504 ref NC_009790.1	227.565	1.16E-57
C153075744_1_225.0:1-223	223	Bacteria	gi 209917191 ref NC_011415.1	425.601	5.87E-117
C153834674_1_302.0:1-300	300	Bacteria	gi 209921952 ref NC_011419.1	558.267	9.37E-157
C151702247_1_148.0:1-146	146	Bacteria	gi 18466424 ref NC_003384.1	277.554	1.33E-72
C151024180_1_126.0:1-124	124	Bacteria	gi 238899406 ref NC_012759.1	235.255	5.91E-60
C152460830_1_183.0:1-181	181	Bacteria	gi 182433793 ref NC_010572.1	173.729	3.07E-41
C153118340_1_228.0:1-226	226	Bacteria	gi 209921952 ref NC_011419.1	425.601	5.95E-117
C153176499_1_233.0:1-231	231	Bacteria	gi 253750923 ref NC_012924.1	415.988	4.78E-114
C150012630_1_105.0:1-103	103	Bacteria	gi 170679574 ref NC_010498.1	192.956	2.53E-47
C152090971_1_164.0:1-162	162	Bacteria	gi 209921952 ref NC_011419.1	302.549	4.49E-80
C152369833_1_178.0:1-176	176	Bacteria	gi 89106884 ref AC_000091.1	321.776	8.06E-86
C153494416_1_263.0:1-261	261	Bacteria	gi 206575367 ref NC_011281.1	292.936	6.00E-77
C151132801_1_129.0:1-127	127	Bacteria	gi 253750923 ref NC_012924.1	223.719	1.81E-56
C150990962_1_126.0:1-124	124	Bacteria	gi 238899406 ref NC_012759.1	214.106	1.37E-53
C152036216_1_162.0:1-160	160	Bacteria	gi 157149574 ref NC_009791.1	237.178	2.11E-60
C156518102_1_1159.0:1-1102	1102	Fungi	gi 160338784 ref NZ_AACM02000411.1	1788.79	0
C154601314_1_425.0:1-423	423	Fungi	gi 160338799 ref NZ_AACM02000426.1	744.768	0
C154321982_1_374.0:1-318	318	Fungi	gi 160338805 ref NZ_AACM02000432.1	606.334	1.63E-171
scaffold32896_1_1686.5:1-273	273	Fungi	gi 160338808 ref NZ_AACM02000435.1	469.823	1.72E-130
scaffold32896_1_1686.2:1-241	241	Fungi	gi 160338808 ref NZ_AACM02000435.1	464.055	8.16E-129
C153540141_1_268.0:1-230	230	Fungi	gi 160338799 ref NZ_AACM02000426.1	402.529	2.57E-110
C153492428_1_263.0:34-237	204	Fungi	gi 160338805 ref NZ_AACM02000432.1	381.38	5.23E-104
C152632064_1_193.0:1-191	191	Fungi	gi 160338794 ref NZ_AACM02000421.1	367.921	5.47E-100
C152529756_1_187.0:1-185	185	Fungi	gi 160338803 ref NZ_AACM02000430.1	344.849	4.65E-93
C152392335_1_179.0:1-177	177	Fungi	gi 160338797 ref NZ_AACM02000424.1	341.003	6.35E-92
C152336937_1_176.0:1-174	174	Fungi	gi 160338803 ref NZ_AACM02000430.1	335.235	3.39E-90
C152085975_1_164.0:1-162	162	Fungi	gi 221228878 ref NZ_AABX02000085.1	312.163	2.78E-83
C152749880_1_201.0:53-199	147	Fungi	gi 160338805 ref NZ_AACM02000432.1	283.322	1.19E-74
C151718669_1_149.0:1-147	147	Fungi	gi 160338797 ref NZ_AACM02000424.1	277.554	6.49E-73
C151503302_1_141.0:1-139	139	Fungi	gi 160338805 ref NZ_AACM02000432.1	267.941	4.76E-70
C151364952_1_136.0:1-134	134	Fungi	gi 160338807 ref NZ_AACM02000434.1	258.328	3.57E-67
C151164119_1_130.0:1-128	128	Fungi	gi 121697430 ref NW_001510354.1	246.791	1.00E-63
C151321645_1_135.0:1-133	133	Fungi	gi 160338790 ref NZ_AACM02000417.1	244.869	3.98E-63
scaffold32896_1_1686.0:1-118	118	Fungi	gi 160338808 ref NZ_AACM02000435.1	227.565	5.57E-58
C151479640_1_140.0:1-138	138	Fungi	gi 160338808 ref NZ_AACM02000435.1	214.106	7.58E-54
C150339437_1_111.0:1-109	109	Fungi	gi 160338805 ref NZ_AACM02000432.1	210.26	8.18E-53
C150340425_1_111.0:1-109	109	Fungi	gi 145606593 ref NW_001798731.1	210.26	8.18E-53
C150285373_1_110.0:1-108	108	Fungi	gi 160338797 ref NZ_AACM02000424.1	208.338	3.06E-52
C150516123_1_114.0:1-112	112	Fungi	gi 145606593 ref NW_001798731.1	200.647	6.62E-50
C152460830_1_183.0:1-181	181	Fungi	gi 115433601 ref XM_001216938.1	198.724	4.41E-49
C149941624_1_104.0:1-102	102	Fungi	gi 145606593 ref NW_001798731.1	196.802	8.46E-49
scaffold39416_1_3596.2:1-506	506	Fungi	gi 221247333 ref NZ_AAYY01000048.1	183.343	5.74E-44
C151256372_1_133.0:1-131	131	Fungi	gi 121697430 ref NW_001510354.1	250.637	7.18E-65
C151702247_1_148.0:1-146	146	Fungi	gi 121697430 ref NW_001510354.1	277.554	6.44E-73
scaffold41546_1_908.0:1-145	145	Fungi	gi 160338807 ref NZ_AACM02000434.1	237.178	9.12E-61

scaffold6349_1_7434.2:211-717	507	Fungi	gi 221247333 ref NZ_AAYY01000048.1	216.028	8.33E-54
C151948881_1_158.0:1-156	156	Fungi	gi 155017865 ref NC_005788.3	194.879	5.38E-48
scaffold60353_1_623.0:100-455	356	Fungi	gi 221247333 ref NZ_AAYY01000048.1	196.802	3.52E-48
scaffold31885_1_3636.3:1-350	350	Fungi	gi 221247333 ref NZ_AAYY01000048.1	216.028	5.64E-54
scaffold3678_491179_492217.1:1-207	207	Fungi	gi 145602701 ref NW_001798719.1	212.183	4.56E-53
scaffold61091_1_2316.1:1-485	485	Fungi	gi 221247333 ref NZ_AAYY01000048.1	177.575	2.99E-42
scaffold76524_1_4387.4:1-513	513	Fungi	gi 221247333 ref NZ_AAYY01000048.1	198.724	1.36E-48
C152369833_1_178.0:1-176	176	Fungi	gi 221228837 ref NZ_AABX02000044.1	308.317	4.36E-82
scaffold31764_1_478.0:1-116	116	Fungi	gi 160338807 ref NZ_AACM02000434.1	202.57	1.83E-50
scaffold11464_1_800.1:1-627	627	Fungi	gi 221247333 ref NZ_AAYY01000048.1	204.492	3.09E-50
scaffold34194_1_9437.2:1-566	566	Fungi	gi 221247333 ref NZ_AAYY01000048.1	221.796	1.71E-55
scaffold42760_1_2579.2:1-331	331	Fungi	gi 221247333 ref NZ_AAYY01000048.1	187.188	2.55E-45
scaffold19674_1_351.0:1-349	349	Fungi	gi 221247333 ref NZ_AAYY01000048.1	194.879	1.31E-47
C151524210_1_142.0:1-140	140	Fungi	gi 121697430 ref NW_001510354.1	239.101	2.31E-61
scaffold21232_1_21043.3:1-533	533	Fungi	gi 221247333 ref NZ_AAYY01000048.1	175.652	1.25E-41
scaffold19121_1_9786.2:1-445	445	Fungi	gi 221247333 ref NZ_AAYY01000048.1	210.26	3.96E-52
scaffold27701_1_2435.2:1-385	385	Fungi	gi 221247333 ref NZ_AAYY01000048.1	183.343	4.31E-44
scaffold42760_1_2579.1:1-276	276	Fungi	gi 221247333 ref NZ_AAYY01000048.1	175.652	6.23E-42
scaffold40930_1_1010.1:1-261	261	Fungi	gi 160338807 ref NZ_AACM02000434.1	385.225	4.79E-105
scaffold8637_1_20241.10:1-597	597	Fungi	gi 221247333 ref NZ_AAYY01000048.1	173.729	5.34E-41
scaffold6349_1_7434.4:1-816	816	Fungi	gi 221247333 ref NZ_AAYY01000048.1	202.57	1.54E-49
scaffold31075_1_3643.1:54-990	937	Fungi	gi 221247333 ref NZ_AAYY01000048.1	242.946	1.24E-61
scaffold21232_1_21043.19:1-649	649	Fungi	gi 221247333 ref NZ_AAYY01000048.1	260.25	5.25E-67

Supplementary Data Set 3. Microbial sequences identified among the ‘no-hit’ contigs from *Wheeler et al.* We aligned the complete set of ‘no hit’ contigs against viral, bacterial, and fungal databases by MegaBlast. The most significant alignments are reported.

Query name	Query length	BLAST Database	Hit ID	Bit score	E-value
Contig142158	206	Viruses	gi 118496614 ref NC_008603.1	362.153	1.30E-99
Contig5230	196	Bacteria	gi 163854304 ref NC_010170.1	279.477	4.93E-73
Contig9647	203	Bacteria	gi 50841496 ref NC_006085.1	362.153	6.65E-98
Contig94873	187	Bacteria	gi 222093774 ref NC_011969.1	262.173	7.57E-68
Contig144673	182	Bacteria	gi 255534169 ref NC_013062.1	183.343	3.94E-44
Contig147766	105	Bacteria	gi 172062142 ref NC_010552.1	100.667	1.57E-19
Contig149702	190	Bacteria	gi 194363778 ref NC_011071.1	244.869	1.25E-62
Contig63162	245	Fungi	gi 145616316 ref NW_001798830.1	187.188	1.84E-45
Contig125249	201	Fungi	gi 145616316 ref NW_001798830.1	175.652	4.38E-42
Contig75100	192	Fungi	gi 145616316 ref NW_001798830.1	179.497	2.89E-43
Contig162560	274	Fungi	gi 145616316 ref NW_001798830.1	225.642	5.52E-57
Contig136490	270	Fungi	gi 145616316 ref NW_001798830.1	204.492	1.26E-50
Contig857	248	Fungi	gi 145601699 ref NW_001798706.1	214.106	1.47E-53
Contig81457	236	Fungi	gi 145616316 ref NW_001798830.1	229.487	3.26E-58
Contig98682	211	Fungi	gi 145616316 ref NW_001798830.1	210.26	1.77E-52
Contig168296	175	Fungi	gi 145616316 ref NW_001798830.1	187.188	1.26E-45
Contig15975	194	Fungi	gi 145615709 ref NW_001798760.1	225.642	3.76E-57
Contig108148	177	Fungi	gi 145616316 ref NW_001798830.1	214.106	1.01E-53
Contig71361	243	Fungi	gi 145616316 ref NW_001798830.1	200.647	1.62E-49
Contig98900	202	Fungi	gi 145616316 ref NW_001798830.1	177.575	1.16E-42
Contig49443	207	Fungi	gi 145616319 ref NW_001798834.1	187.188	1.52E-45
Contig78945	210	Fungi	gi 145616316 ref NW_001798830.1	183.343	2.23E-44
Contig3736	243	Fungi	gi 145616316 ref NW_001798830.1	262.173	4.88E-68
Contig8215	247	Fungi	gi 145616316 ref NW_001798830.1	235.255	6.29E-60
Contig93429	203	Fungi	gi 145616316 ref NW_001798830.1	191.033	1.04E-46
Contig106445	149	Fungi	gi 145616316 ref NW_001798830.1	173.729	1.19E-41
Contig45474	256	Fungi	gi 145616316 ref NW_001798830.1	194.879	9.34E-48
Contig96785	277	Fungi	gi 145615707 ref XM_001414736.1	325.622	4.47E-87
Contig125439	191	Fungi	gi 145616316 ref NW_001798830.1	179.497	2.88E-43
Contig122859	193	Fungi	gi 145616316 ref NW_001798830.1	196.802	1.80E-48
Contig11322	194	Fungi	gi 145616316 ref NW_001798830.1	208.338	6.09E-52
Contig40437	195	Fungi	gi 145616316 ref NW_001798830.1	214.106	1.12E-53
Contig91838	197	Fungi	gi 145616316 ref NW_001798830.1	177.575	1.13E-42
Contig120141	198	Fungi	gi 145616316 ref NW_001798830.1	214.106	1.14E-53
Contig90170	206	Fungi	gi 145616316 ref NW_001798830.1	229.487	2.80E-58
Contig78748	206	Fungi	gi 145601699 ref NW_001798706.1	196.802	1.94E-48
Contig16738	208	Fungi	gi 145616316 ref NW_001798830.1	246.791	1.75E-63
Contig25077	208	Fungi	gi 145616316 ref NW_001798830.1	223.719	1.54E-56
Contig85093	208	Fungi	gi 145601699 ref NW_001798706.1	175.652	4.55E-42
Contig90475	209	Fungi	gi 145616316 ref NW_001798830.1	200.647	1.37E-49
Contig90120	210	Fungi	gi 145616316 ref NW_001798830.1	206.415	2.53E-51