

Supplementary Table 1: GenBank Accession number, genome assembly statistics, and virMine analysis statistics for the 30 metagenomes

Patient	SRA Run Accession #	# of paired reads	# of paired reads after QC	# of singleton reads after QC	# of paired reads mapped to H19	# contigs >1000bp	# bp in contigs >1000bp	# bacterial	# viral	# unknown
C012	ERR926139	477,767	403,786	27,068	33,074	95	164,214	50	4	41
C014	ERR926140	109,093	46,469	12,680	2,655	138	221,567	119	12	7
C019	ERR926141	218,059	169,773	16,673	9,494	55	101,091	22	2	31
C020	ERR926142	429,582	373,893	24,225	34,985	139	1,736,140	133	1	5
C023	ERR926143	444,302	397,258	23,122	29,637	720	1,456,882	666	28	26
C026	ERR926144	399,820	332,890	26,734	36,129	80	134,298	66	9	5
C030	ERR926145	294,961	228,347	21,571	21,268	140	228,170	120	13	7
C032	ERR926146	333,632	298,079	17,015	28,904	57	105,116	23	3	31
C035	ERR926147	354,350	287,819	23,123	17,043	126	1,735,774	119	4	3
C037	ERR926148	228,765	205,830	11,676	18,616	55	96,127	18	5	32
OAB010	ERR926149	1,485,099	1,384,890	74,022	15,099	81	3,996,120	63	8	10
OAB018	ERR926150	1,006,540	947,019	41,013	20,560	409	5,755,177	364	33	12
OAB021	ERR926151	197,665	151,339	15,503	15,891	35	71,730	31	3	1
OAB024	ERR926152	82,502	54,864	7,899	4,010	17	27,226	16	1	0
OAB025	ERR926153	138,664	121,783	8,650	13,501	32	66,041	29	2	1
OAB026	ERR926109	463,425	318,210	36,520	9,414	467	2,384,407	450	12	5
OAB027	ERR926110	299,569	89,654	54,710	4,978	20	33,381	15	3	2
OAB030	ERR926111	278,727	246,089	13,383	20,550	50	97,221	44	3	3
OAB031	ERR926112	287,621	250,033	15,647	22,006	63	108,361	59	2	2
OAB032	ERR926113	309,017	285,263	13,722	14,708	60	1,714,396	52	4	4
OAB039	ERR926114	545,321	461,255	30,878	7,032	279	2,192,535	241	13	25
OAB041	ERR926115	217,193	158,665	17,605	15,506	33	62,683	32	0	1
OAB042	ERR926116	399,050	318,466	28,998	23,764	544	1,216,472	509	11	24
OAB045	ERR926117	585,884	466,727	43,068	23,224	1686	3,579,386	1615	43	28
OAB048	ERR926118	175,272	37,338	30,149	1,185	7	10,888	6	0	1
OAB049	ERR926119	415,230	327,085	20,351	23,290	97	158,079	56	4	37
OAB051	ERR926120	461,737	421,210	24,003	25,779	89	154,556	61	2	26
OAB052	ERR926121	513,060	441,079	29,367	20,687	1352	3,562,632	1276	24	52
OAB056	ERR926122	273,836	250,794	12,151	22,237	59	102,237	52	3	4
OAB062	ERR926123	111,609	29,114	22,329	1,475	11	15,655	1	0	10

Supplementary Table 2: Results for contigs predicted to be viral queried via BLAST against the nr/nt database via the NCBI web interface.

Patient	SRA Run Accession #	Contig #	BLAST hit (Record Description Coding Region Annotation)	Accession #	Contig Length	Identity	Query Coverage
C023	ERR926143	12	NO SIGNIFICANT SIMILARITY (e-value<1)	n/a	7190	n/a	n/a
		57	Homo sapiens BAC clone RP11-1162J9 from chromosome unknown unresolved tandem repeat	AC144610	3994	99%	100%
		68	Human DNA sequence from clone CH507-239L24 on chromosome 21 unannotated	CU638690	3688	87%	100%
		78	Lactobacillus jensenii unannotated	CP018809	3383	100%	1%
		87	Lactobacillus johnsonii transposase	CP021704	3241	74%	44%
		91	Human DNA sequence from clone CH507-239L24 on chromosome 21 unannotated	CU638690	3158	83%	100%
		96	Pan troglodytes BAC clone CH251-408A19 from chromosome 7 unannotated	AC198462	3107	81%	100%
		200	NO SIGNIFICANT SIMILARITY (e-value<1)	n/a	2098	n/a	n/a
		224	Pan troglodytes BAC clone CH251-175J23 from chromosome 7 unannotated	AC188413	1965	100%	1%
		233	Lactobacillus phage Lv-1 DnaC; ssDNA binding protein; hypothetical protein; RusA resolvase	EU871039	1911	83%	99%
		263	Lactobacillus crispatus resolvase	FN692037	1788	92%	29%
		285	NO SIGNIFICANT SIMILARITY (e-value<1)	n/a	1699	n/a	n/a
		288	Lactobacillus jensenii terminase	CP018809	1697	94%	100%
		311	Aerococcus christensenii hypothetical protein; MafB; hypothetical protein	CP014159	1626	78%	54%
		323	Lactobacillus phage Lv-1 minor capsid protein	EU871039	1594	99%	82%
		332	Lactobacillus phage Lv-1 antirepressor	EU871039	1576	82%	37%
		403	Homo sapiens BAC clone CTD-2053H7 from 2 unannotated	AC025223	1414	88%	100%
		408	Streptococcus pneumoniae putative membrane protein; hypothetical protein	HG799494	1401	97%	100%
		420	NO SIGNIFICANT SIMILARITY (e-value<1)	n/a	1387	n/a	n/a
		446	NO SIGNIFICANT SIMILARITY (e-value<1)	n/a	1355	n/a	n/a
		460	NO SIGNIFICANT SIMILARITY (e-value<1)	n/a	1330	n/a	n/a

Supplementary Table 2: Results for contigs predicted to be viral queried via BLAST against the nr/nt database via the NCBI web interface.

Patient	SRA Run Accession #	Contig #	BLAST hit (Record Description Coding Region Annotation)	Accession #	Contig Length	Identity	Query Coverage
C023	ERR926143	466	NO SIGNIFICANT SIMILARITY (e-value<1)	n/a	1314	n/a	n/a
		483	Lactobacillus jensenii tRNA-Ser	CP018809	1288	100%	3%
		506	Lactobacillus jensenii glycosyl hydrolase family 25	CP018809	1249	96%	50%
		523	NO SIGNIFICANT SIMILARITY (e-value<1)	n/a	1227	n/a	n/a
		608	Lactobacillus phage Lv-1 terminase large subunit; portal protein	EU871039	1113	94%	99%
		688	Homo sapiens FOSMID clone ABC12-47837900C21 from chromosome 9 unresolved duplication	AC242285	1034	98%	100%
		698	NO SIGNIFICANT SIMILARITY (e-value<1)	n/a	1022	n/a	n/a
C026	ERR926144	15	Homo sapiens FOSMID clone ABC12-47837900C21 from chromosome 9 unresolved duplication	AC242285	2140	98%	100%
		23	NO SIGNIFICANT SIMILARITY (e-value<1)	n/a	1867	n/a	n/a
		27	Homo sapiens chromosome 1 clone RP11-504P24 unannotated	AC138393	1803	75%	99%
		36	Corynebacterium striatum transposase	CP024932	1541	99%	98%
		57	Homo sapiens BAC clone CH17-116G12 from chromosome 2 unresolved tandem repeat	AC255551	1250	78%	98%
		63	Homo sapiens FOSMID clone ABC16-1709J7 from chromosome 4 unannotated	AC241532	1140	80%	99%
		69	Human (clone TRI-6) satellite I repeat region repeat	L01057	1109	86%	100%
		72	NO SIGNIFICANT SIMILARITY (e-value<1)	n/a	1071	n/a	n/a
		80	Spirometra erinaceieuropaei unannotated	LN146799	1003	96%	57%
C030	ERR926145	9	Lactobacillus crispatus ST1 lysin; membrane protein; penicillin-binding protein	FN692037	2838	99%	100%
		10	Homo sapiens BAC clone RP11-1162J9 from chromosome unknown unannotated	AC144610	2823	99%	100%
		14	Lactobacillus crispatus ST1 transposase	FN692037	2380	98%	60%
		15	Lactobacillus amylovorus transposase	CP002338	2293	100%	91%
		39	Lactobacillus helveticus transposase	CP002429	1498	98%	80%

Supplementary Table 2: Results for contigs predicted to be viral queried via BLAST against the nr/nt database via the NCBI web interface.

Patient	SRA Run Accession #	Contig #	BLAST hit (Record Description Coding Region Annotation)	Accession #	Contig Length	Identity	Query Coverage
OAB018	ERR926150	28	Lactobacillus helveticus hypothetical protein; minor capsid protein; capsid; terminase; terminase; phage tail protein	CP015496	37166	87%	71%
		49	Lactobacillus helveticus lyson; phage holin; hypothetical protein; hypothetical protein	CP015496	26779	85%	15%
		66	NO SIGNIFICANT SIMILARITY (e-value<1)	n/a	17753	n/a	n/a
			Lactobacillus allii phage tail tape measure protein	CP019323	17753	72%	3%
		125	NO SIGNIFICANT SIMILARITY (e-value<1)	n/a	9400	n/a	n/a
		148	Lactobacillus helveticus hypothetical protein; capsid protein	CP011386	7640	76%	25%
		153	NO SIGNIFICANT SIMILARITY (e-value<1)	n/a	7248	n/a	n/a
			Lactobacillus helveticus hypothetical proteins	CP011386	7248	70%	95%
		175	Lactobacillus gallinarum transposase	CP012890	5965	91%	29%
		176	Lactobacillus helveticus hypothetical protein; hypothetical protein	CP015496	5964	83%	4%
		198	Lactobacillus helveticus terminase; restriction endonuclease	CP011386	5099	89%	53%
		207	NO SIGNIFICANT SIMILARITY (e-value<1)	n/a	4704	n/a	n/a
		208	Lactobacillus crispatus conserved protein; rRNA	FN692037	4655	99%	8%
		213	Lactobacillus crispatus conserved protein; hypothetical protein; LexA repressor; Acyl-coa thioesterase	FN692037	4583	97%	99%
		214	Lactobacillus amylovorus predicted membrane protein; transposase	CP002338	4575	93%	38%
		217	NO SIGNIFICANT SIMILARITY (e-value<1)	n/a	4372	n/a	n/a
		226	NO SIGNIFICANT SIMILARITY (e-value<1)	n/a	4120	n/a	n/a
		269	Lactobacillus crispatus tRNAs	FN692037	2954	99%	5%
		275	Lactobacillus helveticus DUF2479 domain-containing protein	CP015496	2826	84%	27%
		284	NO SIGNIFICANT SIMILARITY (e-value<1)	n/a	2579	n/a	n/a
		300	NO SIGNIFICANT SIMILARITY (e-value<1)	n/a	2291	n/a	n/a
		312	Lactobacillus helveticus hypothetical protein	CP015496	2073	81%	36%
		319	Lactobacillus crispatus ATP-dependent helicase/nuclease subunit A	FN692037	2019	99%	100%
		323	Homo sapiens BAC clone RP11-237A9 from 4 repeat	AC129664	1966	90%	99%
		332	Lactobacillus helveticus insertion sequence	CP002081	1862	98%	62%
		338	NO SIGNIFICANT SIMILARITY (e-value<1)	n/a	1812	n/a	n/a

Supplementary Table 2: Results for contigs predicted to be viral queried via BLAST against the nr/nt database via the NCBI web interface.

Patient	SRA Run Accession #	Contig #	BLAST hit (Record Description Coding Region Annotation)	Accession #	Contig Length	Identity	Query Coverage
OAB018	ERR926150	348	Lactobacillus helveticus lysin	CP015496	1689	73%	43%
		352	Lactobacillus jensenii hypothetical protein	CP018809	1638	81%	38%
		366	Lactobacillus crispatus transposase	FN692037	1425	99%	100%
		375	Human DNA sequence from clone RP13-511L2 on chromosome X unannotated	BX537339	1294	85%	100%
		377	NO SIGNIFICANT SIMILARITY (e-value<1)	n/a	1280	n/a	n/a
		379	NO SIGNIFICANT SIMILARITY (e-value<1)	n/a	1248	n/a	n/a
		389	Pan troglodytes BAC clone CH251-398H5 from chromosome 7 unannotated	AC217674	1162	95%	100%
		396	Lactobacillus crispatus conserved protein; Nicotinamide mononucleotide transporter	FN692037	1115	99%	100%
		403	NO SIGNIFICANT SIMILARITY (e-value<1)	n/a	1088	n/a	n/a
OAB021	ERR926151	1	JC virus variant GS/K complete genome	AF004349	5219	99%	100%
		17	Human (clone TRI-6) satellite I repeat region tandem repeat	L01057	1734	83%	99%
		29	Homo sapiens gene alpha satellite in centromeric region of chromosome 21 alpha satellite	D29750	1289	99%	100%
OAB024	ERR926152	10	Human (clone TRI-6) satellite I repeat region tandem repeat	L01057	1427	92%	100%
OAB025	ERR926153	30	NO SIGNIFICANT SIMILARITY (e-value<1)	n/a	1143	n/a	n/a
		32	JC virus variant GS/K agno; VP2	AF004349	1023	99%	100%
OAB026	ERR926109	22	JC virus variant GS/K complete genome	AF004349	5220	98%	100%
		152	Gardnerella vaginalis translation initiation factor IF-2	CP002725	1950	99%	100%
		180	Human DNA sequence from clone CH507-416A10 on chromosome 21 unannotated	FP236243	1822	98%	100%
		182	Human DNA sequence from clone RP11-145C1 unannotated	AL162451	1817	86%	100%
		226	Gardnerella vaginalis hypothetical protein; repeat-containing protein	LT629773	1581	99%	100%

Supplementary Table 2: Results for contigs predicted to be viral queried via BLAST against the nr/nt database via the NCBI web interface.

Patient	SRA Run Accession #	Contig #	BLAST hit (Record Description Coding Region Annotation)	Accession #	Contig Length	Identity	Query Coverage
OAB032	ERR926113	18	JC polyomavirus DNA isolate UZ-20 complete genome	AB127023	5220	99%	100%
		23	Homo sapiens clone fosmid ABC8_4_1_000041143400_G2 unannotated	GQ222231	2496	90%	99%
		42	Human DNA sequence from clone CH507-416A10 on chromosome 21 unannotated	FP236243	1570	99%	100%
		56	Homo sapiens BAC clone RP13-489M19 from 4 unannotated	AC127385	1030	99%	100%
OAB039	ERR926114	55	Lactobacillus amylovorus unannotated	CP002338	13596	100%	1%
			Lactobacillus allii phage tail tape measure protein	CP019323	13596	72%	4%
		79	Lactobacillus helveticus unannotated	CP015444	8504	100%	1%
			Lactobacillus bacteriophage phi adh	AJ131519	8504	67%	54%
		85	Lactobacillus amylovorus transposase	CP017706	7741	99%	2%
		111	Lactobacillus johnsonii polar amino acid ABC transporter permease	CP021704	5691	100%	2%
		118	Lactobacillus crispatus peptide methionine sulfoxide reductase	FN692037	5195	99%	35%
		121	NO SIGNIFICANT SIMILARITY (e-value<1)	n/a	4703	n/a	n/a
		151	Lactobacillus amylovorus transposase	CP017706	3186	98%	2%
		197	Actinotignum schaalii hypothetical protein	CP008802	1795	92%	16%
		214	Lactobacillus amylovorus unannotated	CP002338	1540	98%	5%
		220	Lactobacillus crispatus tRNA-Ser	FN692037	1423	100%	30%
		230	Streptococcus anginosus hypothetical protein; predicted membrane protein	AP013072	1329	94%	88%
		263	Lactobacillus crispatus conserved protein; Nicotinamide mononucleotide transporter	FN692037	1082	99%	100%
		264	NO SIGNIFICANT SIMILARITY (e-value<1)	n/a	1075	n/a	n/a
OAB041	ERR926115	NO VIRAL CONTIGS IDENTIFIED					

Supplementary Table 2: Results for contigs predicted to be viral queried via BLAST against the nr/nt database via the NCBI web interface.

Patient	SRA Run Accession #	Contig #	BLAST hit (Record Description Coding Region Annotation)	Accession #	Contig Length	Identity	Query Coverage
OAB042	ERR926116	12	Homo sapiens FOSMID clone ABC13-1016922M5 from chromosome 9 unresolved tandem repeat	AC244148	6649	92%	100%
		26	JC virus variant GS/K complete genome	AF004349	5220	98%	100%
		116	Homo sapiens chromosome 16 clone CTD-2506I5 unannotated	AC137488	2788	94%	100%
		118	Homo sapiens chromosome 21 satellite I sequence satellite	JX174276	2767	84%	100%
		256	Gardnerella vaginalis hypothetical protein	LT629773	1827	77%	63%
		323	Gardnerella vaginalis inosine-5'-monophosphate dehydrogenase	CP002725	1544	92%	99%
		371	NO SIGNIFICANT SIMILARITY (e-value<1)	n/a	1405	n/a	n/a
		398	Gardnerella vaginalis repeat protein	CP002725	1335	81%	99%
		501	Homo sapiens BAC clone RP11-18C16 from 4 CpG island	AC110072	1096	81%	98%
		504	Gardnerella vaginalis hypothetical protein	CP002725	1088	84%	57%
		519	Gardnerella vaginalis chromosomal replication initiator protein	CP002725	1050	85%	92%
OAB045	ERR926117	87	JC polyomavirus strain #20 complete genome	MF662204	5220	99%	100%
		144	Homo sapiens BAC clone RP11-1162J9 from chromosome unknown unresolved tandem repeat	AC144610	4072	99%	100%
		203	Homo sapiens FOSMID clone ABC13-1016922M5 from chromosome 9 unresolved tandem repeat	AC244148	3515	99%	100%
		327	Enterobacter aerogenes viral enhancin protein; hypothetical protein; 2,3,4,5-tetrahydropyridine-2,6-carboxylate N-succinyltransferase	CP002824	2664	99%	100%
		371	Klebsiella aerogenes hypothetical protein; lysozyme	CP011539	2489	93%	78%
		456	Klebsiella aerogenes phage tail protein; phage tail protein	CP014747	2208	90%	82%
		464	Gardnerella vaginalis hypothetical protein; hypothetical protein	CP002725	2188	97%	100%
		496	Enterobacter aerogenes putative transporter	CP002824	2079	99%	43%
		548	Klebsiella aerogenes hypothetical protein; host specificity protein J	CP011539	1963	95%	99%
		550	Klebsiella aerogenes bifunctional methylenetetrahydrofolate dehydrogenase; hypothetical protein	CP023963	1961	99%	100%
		555	Gardnerella vaginalis peptide/nickel transport system substrate-binding protein; pyridoxal 5'-phosphate synthase pdxS subunit	LT629773	1945	99%	100%

Supplementary Table 2: Results for contigs predicted to be viral queried via BLAST against the nr/nt database via the NCBI web interface.

Patient	SRA Run Accession #	Contig #	BLAST hit (Record Description Coding Region Annotation)	Accession #	Contig Length	Identity	Query Coverage
OAB045	ERR926117	564	Homo sapiens BAC clone RP13-489M19 from 4 unresolved tandem repeat	AC127385	1921	99%	100%
		707	Klebsiella aerogenes hypothetical protein; P22 coat protein	CP024883	1726	99%	100%
		789	Homo sapiens BAC clone RP11-886I11 from Y L1 repeat	AC134882	1607	88%	100%
		798	Klebsiella aerogenes phage tail tape measure protein	CP014747	1597	97%	100%
		867	Gardnerella vaginalis alkyl hydroperoxide reductase subunit F; peroxiredoxin (alkyl hydroperoxide reductase subunit C)	LT629773	1518	96%	100%
		892	Klebsiella aerogenes long-chain fatty acid--CoA ligase; DNA methyltransferase	CP014748	1496	96%	100%
		900	Gardnerella vaginalis ribosomal proteins	CP002725	1486	100%	100%
		969	Gardnerella vaginalis putative lipoprotein	CP002725	1410	99%	100%
		970	Klebsiella aerogenes glutathione S-transferase; capsular biosynthesis protein	CP024880	1410	99%	100%
		991	Gardnerella vaginalis hypothetical protein; signal transduction histidine kinase	CP002725	1394	100%	100%
		994	Pan troglodytes BAC clone CH251-398H5 from chromosome 7 unresolved tandem repeat	AC217674	1389	97%	99%
		1037	Klebsiella aerogenes terminase; hypothetical protein	CP014747	1358	99%	100%
		1053	Klebsiella aerogenes primosomal protein 1; regulatory protein	CP011539	1347	98%	98%
		1080	Gardnerella vaginalis phosphoribosyl transferase domain protein; ATPase/histidine kinase/DNA gyrase B/HSP90 domain protein	CP002725	1324	99%	100%
		1094	Gardnerella vaginalis uncharacterized OsmC-related protein; thymidylate synthase	LT629773	1314	99%	100%
		1157	Klebsiella aerogenes Hsp33 family molecular chaperone; phosphoenolpyruvate carboxykinase	CP024883	1274	99%	100%
		1170	Raoultella ornithinolytica capsid protein; serine peptidase	CP012555	1268	91%	100%
		1194	Bacteriophage phiKO2 putative portal protein; putative head maturation protease	AY374448	1252	95%	100%

Supplementary Table 2: Results for contigs predicted to be viral queried via BLAST against the nr/nt database via the NCBI web interface.

Patient	SRA Run Accession #	Contig #	BLAST hit (Record Description Coding Region Annotation)	Accession #	Contig Length	Identity	Query Coverage	
OAB045	ERR926117	1259	Klebsiella aerogenes GNAT family N-acetyltransferase; hypothetical protein	CP024885	1207	99%	76%	
		1321	Homo sapiens BAC clone RP11-108I14 from Y centromeric repeat	AC010970	1170	87%	100%	
		1334	Gardnerella vaginalis DNA binding domain-containing protein, excisionase family; hypothetical protein	LT629773	1162	99%	100%	
		1338	Homo sapiens chromosome 21 satellite I sequence satellite	JX174276	1160	84%	99%	
		1361	Homo sapiens chromosome 21 satellite I sequence satellite	JX174276	1148	99%	99%	
		1398	Klebsiella aerogenes terminase	CP024883	1136	99%	99%	
		1424	Klebsiella aerogenes phage tail tape measure protein	CP014747	1124	98%	100%	
		1482	Gardnerella vaginalis ABC-2 type transport system ATP-binding protein	LT629773	1098	97%	100%	
		1510	Enterobacter aerogenes hemagglutinin domain-containing protein; putative entero exodeoxyribonuclease VIII	CP002824	1082	97%	100%	
		1552	Klebsiella aerogenes myelin regulatory factor	CP011539	1057	80%	54%	
		1580	Klebsiella aerogenes GNAT family N-acetyltransferase; cysteine tRNA ligase	CP024885	1044	99%	100%	
		1582	Gardnerella vaginalis alpha-mannosidase	LT629773	1044	93%	94%	
		1654	Klebsiella pneumoniae virulence-related outer membrane protein	AP014950	1013	85%	38%	
		1657	Klebsiella aerogenes lysine decarboxylase LdcC; lysine:cadaverine antiporter	CP024883	1012	99%	100%	
OAB048	ERR926118	NO VIRAL CONTIGS IDENTIFIED						
OAB049	ERR926119	44	NO SIGNIFICANT SIMILARITY (e-value<1)	n/a	1427	n/a	n/a	
		46	NO SIGNIFICANT SIMILARITY (e-value<1)	n/a	1392	n/a	n/a	
		62	Lactobacillus salivarius transposase	CP024065	1210	90%	95%	
		91	NO SIGNIFICANT SIMILARITY (e-value<1)	n/a	1040	n/a	n/a	

Supplementary Table 2: Results for contigs predicted to be viral queried via BLAST against the nr/nt database via the NCBI web interface.

Patient	SRA Run Accession #	Contig #	BLAST hit (Record Description Coding Region Annotation)	Accession #	Contig Length	Identity	Query Coverage
OAB051	ERR926120	75	Human DNA sequence from clone CH17-294F12 on chromosome 20 unannotated	FO681521	1031	77%	99%
		86	Corynebacterium jeikeium strain FDAARGOS_328 transposase	CP022054	1007	100%	100%
OAB052	ERR926121	139	Gardnerella vaginalis phage terminase; phage portal protein; endopeptidase Clp; major capsid protein; phage DNA packaging protein; hypothetical protein; hypothetical protein	CP002725	4898	96%	99%
		330	Gardnerella vaginalis hypothetical protein	CP019058	3131	99%	100%
		361	Gardnerella vaginalis phage antirepressor; hypothetical protein	CP019058	2958	94%	20%
		368	NO SIGNIFICANT SIMILARITY (e-value<1)	n/a	2941	n/a	n/a
		435	uncultured bacterium clone transposase	KU546379	2582	92%	39%
		737	uncultured bacterium clone endonuclease	KU545938	1719	97%	30%
		782	Prevotella melaninogenica transposase	CP022040	1654	77%	93%
		802	Gardnerella vaginalis hypothetical protein; hypothetical protein; hypothetical protein; putative DNA-directed DNA polymerase	CP002725	1626	96%	99%
		825	Lactobacillus gasseri putative transposase	CP006803	1578	99%	100%
		828	Homo sapiens chromosome 21 satellite I sequence satellite	JX174276	1573	84%	100%
		892	Gardnerella vaginalis hypothetical protein	CP019058	1474	98%	100%
		906	NO SIGNIFICANT SIMILARITY (e-value<1)	n/a	1459	n/a	n/a
		948	Streptococcus pyogenes conjugal transfer protein; antirestriction protein ArdA; hypothetical protein; replication initiation protein		1397	99%	100%
973	NO SIGNIFICANT SIMILARITY (e-value<1)	n/a	1367	n/a	n/a		
999	NO SIGNIFICANT SIMILARITY (e-value<1)	n/a	1340	n/a	n/a		
1070	Gardnerella vaginalis hypothetical protein; Ca-activated chloride channel family protein	LT629773	1270	78%	96%		
1072	Lactobacillus jensenii unannotated region	CP018809	1267	98%	12%		
1151	NO SIGNIFICANT SIMILARITY (e-value<1)	n/a	1178	n/a	n/a		
1186	Homo sapiens homeobox protein DUX3 (DUX3) gene double homeobox protein	AF133130	1141	90%	100%		

Supplementary Table 2: Results for contigs predicted to be viral queried via BLAST against the nr/nt database via the NCBI web interface.

Patient	SRA Run Accession #	Contig #	BLAST hit (Record Description Coding Region Annotation)	Accession #	Contig Length	Identity	Query Coverage	
OAB052	ERR926121	1230	Gardnerella vaginalis acetate kinase	LT629773	1095	97%	100%	
		1233	Gardnerella vaginalis hypothetical protein	CP002725	1092	93%	92%	
		1259	Gardnerella vaginalis DNA repair protein RadA	CP002725	1068	84%	99%	
		1269	Aerococcus christensenii DNA-binding protein	CP014159	1060	97%	100%	
		1281	Streptococcus anginosus hypothetical protein; hypothetical protein	AP013072	1050	89%	100%	
OAB056	ERR926122	23	NO SIGNIFICANT SIMILARITY (e-value<1)	n/a	1583	n/a	n/a	
		36	Human DNA sequence from clone CH507-39O4 on chromosome 21	FP236315	1319	99%	100%	
		48	Homo sapiens BAC clone RP11-108I14 from Y centromeric repeat	AC010970	1116	87%	92%	
OAB062	ERR926123	NO VIRAL CONTIGS IDENTIFIED						

**Supplementary Table 3: Annotations for putative complete/
near-complete phage genomes listed in Table 1.**

OAB010 - contig 28			
start	stop	strand	function
112	2787	+	Phage tail length tape-measure protein
2784	3509	+	FIG01115816: hypothetical protein
3509	6424	+	Phage endopeptidase
6437	8293	+	Phage structural protein
8311	8709	+	Holin, toxin secretion/phage lysis
8711	10117	+	Phage-associated cell wall hydrolase
10898	12139	+	Recombinase
12102	13388	+	Site-specific recombinase
13458	14231	+	FIG01120285: hypothetical protein
14245	15750	+	FIG01110938: hypothetical protein
16000	16212	+	hypothetical CDS
16330	17067	+	Class B acid phosphatase precursor (EC 3.1.3.2)
17439	17305	-	hypothetical protein
OAB010 - contig 31			
start	stop	strand	function
2467	1742	-	FIG01115816: hypothetical protein
5583	2464	-	Phage tail length tape-measure protein
6297	5878	-	prophage pi2 protein 40
6878	6309	-	Prophage pi2 protein 39
7207	6881	-	prophage pi2 protein 38
7584	7216	-	prophage pi2 protein 37
7915	7577	-	FIG01117178: hypothetical protein
8097	7915	-	unknown phage protein
OAB010 - contig 39			
start	stop	strand	function
901	482	-	prophage pi2 protein 40
1482	913	-	Prophage pi2 protein 39
1811	1485	-	prophage pi2 protein 38
2188	1820	-	prophage pi2 protein 37
2519	2181	-	FIG01117178: hypothetical protein
2776	2519	-	unknown phage protein
3453	2773	-	Phage major capsid protein

**Supplementary Table 3: Annotations for putative complete/
near-complete phage genomes listed in Table 1.**

OAB018 - contig 28

start	stop	strand	function
4008	901	-	DNA polymerase III alpha subunit (EC 2.7.7.7)
4124	4342	+	FIG00743751: hypothetical protein
5960	4404	-	2',3'-cyclic-nucleotide 2'-phosphodiesterase (EC 3.1.4.16)
6255	6076	-	FIG00752837: hypothetical protein
6562	6443	-	hypothetical protein
7613	6726	-	Lysin
8058	7603	-	Phage protein
8299	8048	-	hypothetical protein
8731	8312	-	hypothetical protein
8855	8694	-	hypothetical protein
9541	8855	-	hypothetical protein
9873	9586	-	hypothetical protein
10525	10079	-	hypothetical protein
10993	10583	-	hypothetical protein
12640	11006	-	hypothetical protein
12833	12600	-	hypothetical protein
16288	12842	-	Phage endopeptidase
17070	16288	-	Phage-related protein
23815	17060	-	Phage tail length tape-measure protein
24459	23815	-	Phage protein
24901	24467	-	FIG00743450: hypothetical protein
25604	24924	-	major capsid protein gpP
26039	25605	-	minor capsid protein
26428	26039	-	minor capsid protein
26778	26428	-	Phage capsid and scaffold
27163	26771	-	FIG00754064: hypothetical protein
28246	27179	-	minor capsid protein
28876	28256	-	Phage capsid scaffolding protein
29190	28972	-	hypothetical protein
30325	29192	-	Phage minor capsid protein
31934	30309	-	minor capsid protein
33202	31946	-	Terminase large subunit [Bacteriophage A118]
33893	33219	-	FIG00746678: hypothetical protein
34607	33903	-	ParB domain protein nuclease
34957	34610	-	hypothetical protein
35525	35082	-	hypothetical protein
36124	35525	-	hypothetical protein
36290	36144	-	hypothetical protein
36657	36319	-	FIG00748422: hypothetical protein

**Supplementary Table 3: Annotations for putative complete/
near-complete phage genomes listed in Table 1.**

OAB018 - contig 49			
start	stop	strand	function
640	254	-	FIG00752284: hypothetical protein
1539	799	-	putative phage minor head protein
2992	1589	-	FIG00748140: hypothetical protein
4459	3008	-	Phage terminase, large subunit
4997	4452	-	Phage Terminase Small Subunit
5096	5275	+	hypothetical protein
5325	5717	+	Phage-related protein
6450	6094	-	hypothetical protein
6623	6438	-	FIG00752030: hypothetical protein
7079	6624	-	FIG00742225: hypothetical protein
7382	7248	-	hypothetical protein
9507	8983	-	Phage related protein
10011	9526	-	Phage Holliday junction resolvase
10421	10008	-	hypothetical protein
10714	10430	-	hypothetical protein
11102	10701	-	hypothetical protein
11547	11311	-	hypothetical protein
11869	11534	-	hypothetical protein
12670	11873	-	Helicase loader Dnal
13488	12682	-	putative phage replication protein
14267	13491	-	Phage protein
14517	14260	-	hypothetical protein
14826	14527	-	hypothetical protein
15053	14880	-	hypothetical protein
15291	15058	-	hypothetical protein
15615	15301	-	hypothetical protein
16515	15631	-	Phage antirepressor protein
16740	16531	-	hypothetical protein
16891	17229	+	CI phage repressor protein
17241	17642	+	hypothetical protein
17645	18010	+	Lj928 prophage protein
18023	18523	+	hypothetical protein
18623	19753	+	Integrase
21694	21488	-	hypothetical protein
21916	21728	-	hypothetical protein
22221	22099	-	hypothetical protein
23318	22395	-	Lysin
23763	23308	-	Phage protein
23957	23760	-	hypothetical protein
24365	23970	-	hypothetical protein
OAB018 - contig 49 (cont)			

**Supplementary Table 3: Annotations for putative complete/
near-complete phage genomes listed in Table 1.**

24591	24328	-	hypothetical protein
25148	24591	-	hypothetical protein
25506	25231	-	hypothetical protein
26137	25496	-	hypothetical protein
26560	26138	-	hypothetical protein
OAB018 - contig 66			
start	stop	strand	function
50	439	+	hypothetical protein
429	758	+	FIG00744673: hypothetical protein
751	1149	+	Phage tail assembly
1146	1520	+	putative tail component
1520	2296	+	Phage major tail protein
2377	2790	+	hypothetical protein
2841	2990	+	hypothetical protein
3024	10043	+	Phage tail length tape-measure protein
10033	10794	+	FIG00743423: hypothetical protein
10764	11825	+	FIG00744650: hypothetical protein
11815	12561	+	FIG00744650: hypothetical protein
12576	14186	+	FIG00744650: hypothetical protein
14191	14385	+	hypothetical protein
14357	14740	+	hypothetical protein
14794	17181	+	hypothetical protein
17198	17593	+	FIG00748181: hypothetical protein
OAB018 - contig 148			
start	stop	strand	function
1148	996	-	hypothetical protein
1557	1210	-	hypothetical protein
2361	1651	-	major tail protein
2729	2343	-	Phage tail protein
3163	2726	-	Phage tail assembly
3467	3156	-	hypothetical protein
3819	3478	-	hypothetical protein
5070	3838	-	Phage capsid protein
5843	5082	-	Prophage Clp protease-like protein
6999	5824	-	Phage portal protein
7193	6999	-	hypothetical protein
7440	7186	-	Phage terminase, large subunit

**Supplementary Table 3: Annotations for putative complete/
near-complete phage genomes listed in Table 1.**

OAB039 - contig 55			
start	stop	strand	function
480	97	-	hypothetical protein
646	452	-	hypothetical protein
4058	651	-	FIG00744650: hypothetical protein
4819	4058	-	FIG00743423: hypothetical protein
11828	4809	-	Phage tail length tape-measure protein
12011	11862	-	hypothetical protein
12475	12062	-	hypothetical protein
13332	12556	-	Phage major tail protein
13550	13332	-	putative tail component
OAB039 - contig 79			
start	stop	strand	function
270	43	-	putative tail component
665	267	-	Phage tail assembly
987	658	-	FIG00744673: hypothetical protein
1366	977	-	hypothetical protein
2741	1383	-	Phage capsid protein
3383	2760	-	Prophage Clp protease-like protein
4587	3409	-	Phage portal protein
4802	4590	-	hypothetical protein
6687	4795	-	Phage terminase, large subunit
6872	6708	-	hypothetical protein
7071	6931	-	hypothetical protein
7490	7068	-	Phage terminase small subunit
8129	7650	-	Phage-associated homing endonuclease
8281	8168	-	hypothetical protein

Supplementary Figure 1: Workflow for bioinformatic analysis.

