

Supplemental Material for

Metagenomic characterization of airborne viral DNA diversity in the near-surface atmosphere

Tae Woong Whon, Min-Soo Kim, Seong Woon Roh, Na-Ri Shin, Hae-Won Lee, and Jin-
Woo Bae*

*To whom correspondence should be addressed. E-mail: baejw@khu.ac.kr

This PDF file includes:

FIG. S1 to S4

TABLES S1 to S4

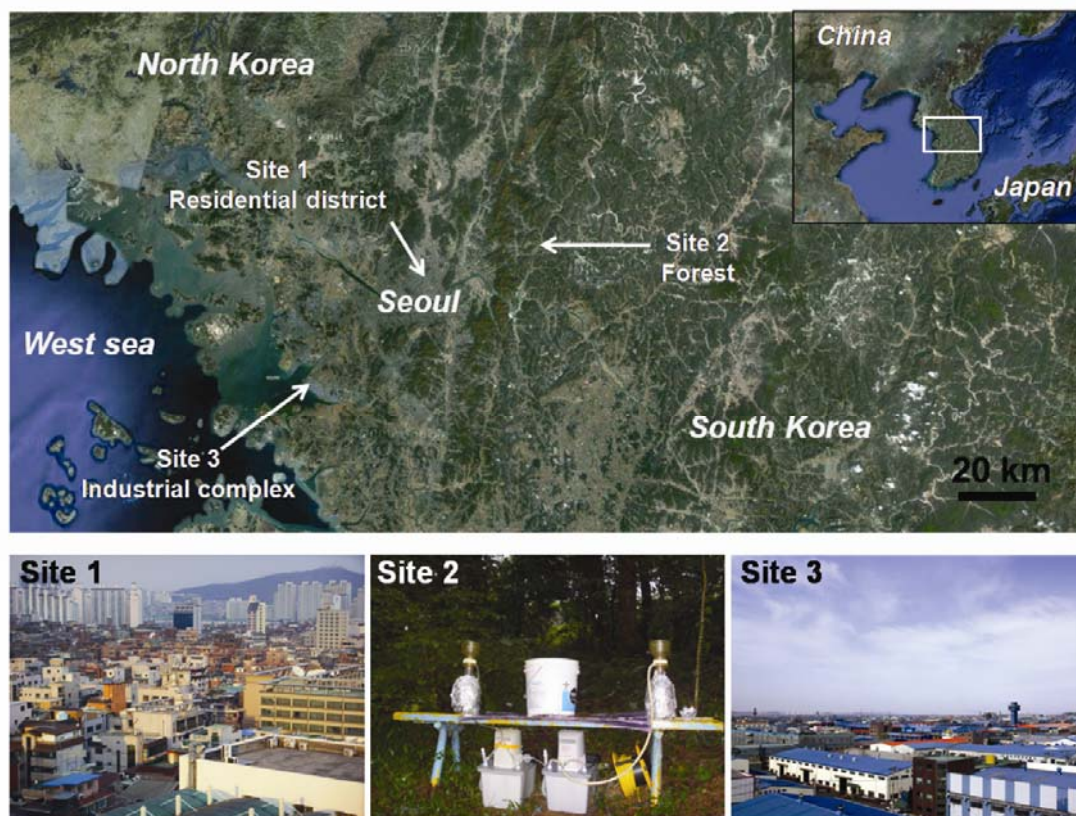


FIG. S1. Sampling sites (map from <http://earth.google.com/>) and land-use type images. Site 1 (RD) was located in the middle of Seoul city. Sites 2 (FR) and 3 (IC) were located in the central region and west coast of Korea, respectively.

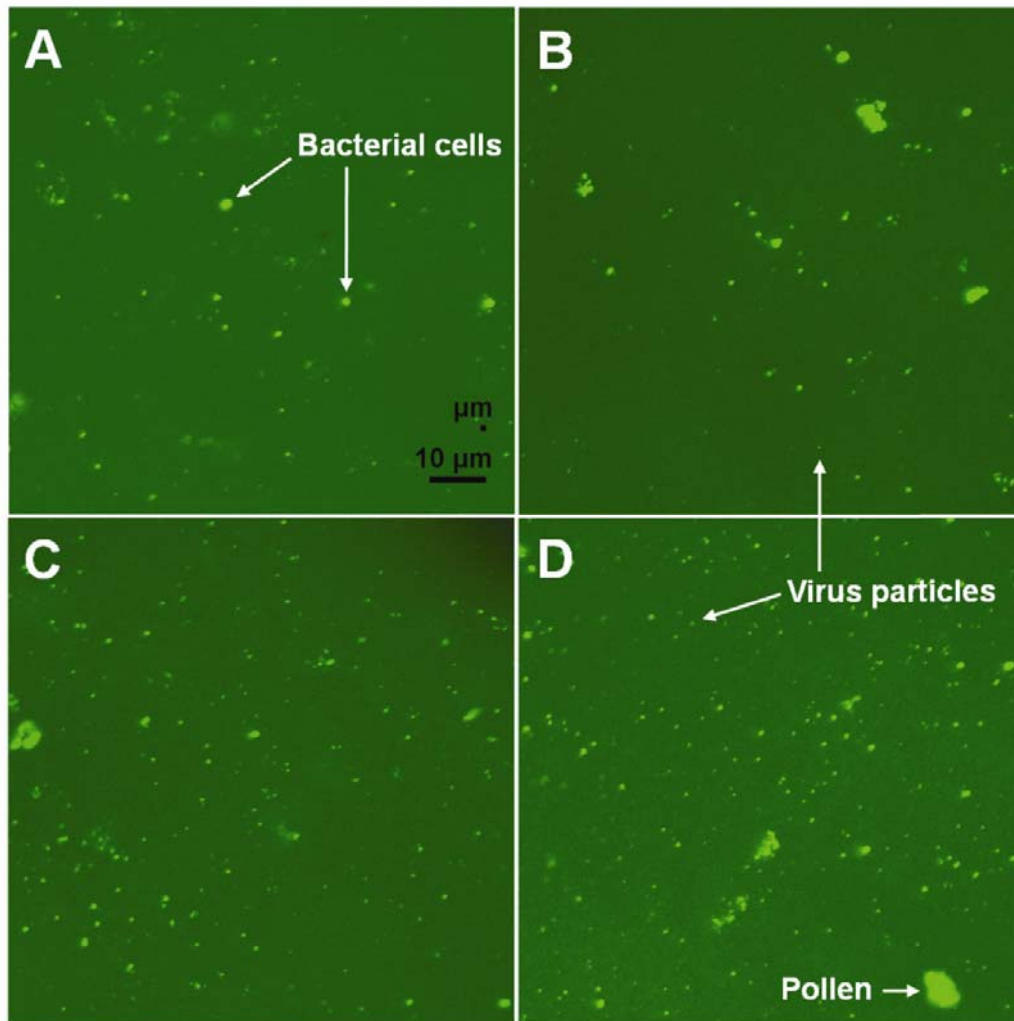


FIG. S2. Virus particles stained with SYBR-Gold and visualized under confocal laser scanning microscopy. Airborne particles containing viruses and bacteria were collected from RD during October (A), November (B), December, 2010 (C), and January, 2011 (D).

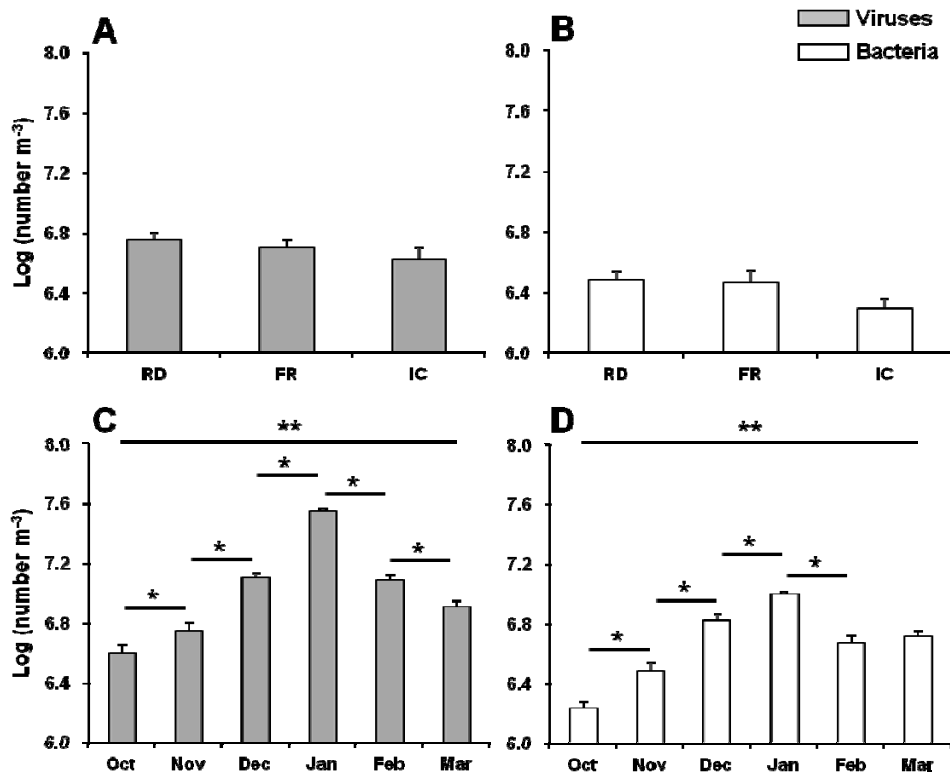


FIG. S3. Spatial and temporal abundance of airborne viruses and bacteria collected from the three different land-use types. For spatial variation, total viral (A, $P = 0.228$, one-way ANOVA) and bacterial (B, $P = 0.136$, one-way ANOVA) abundance in the samples collected from the residential district (RD), forest (FR), and industrial complex (IC) on November, 2010. For temporal variation, total viral (C) and bacterial (D) abundance in the samples collected at the RD site during 6 months. For C and D, $n = 8-10$; * $P < 0.05$, ** $P < 0.001$, one-way ANOVA with Duncan correction. Error bars represent standard error of means.

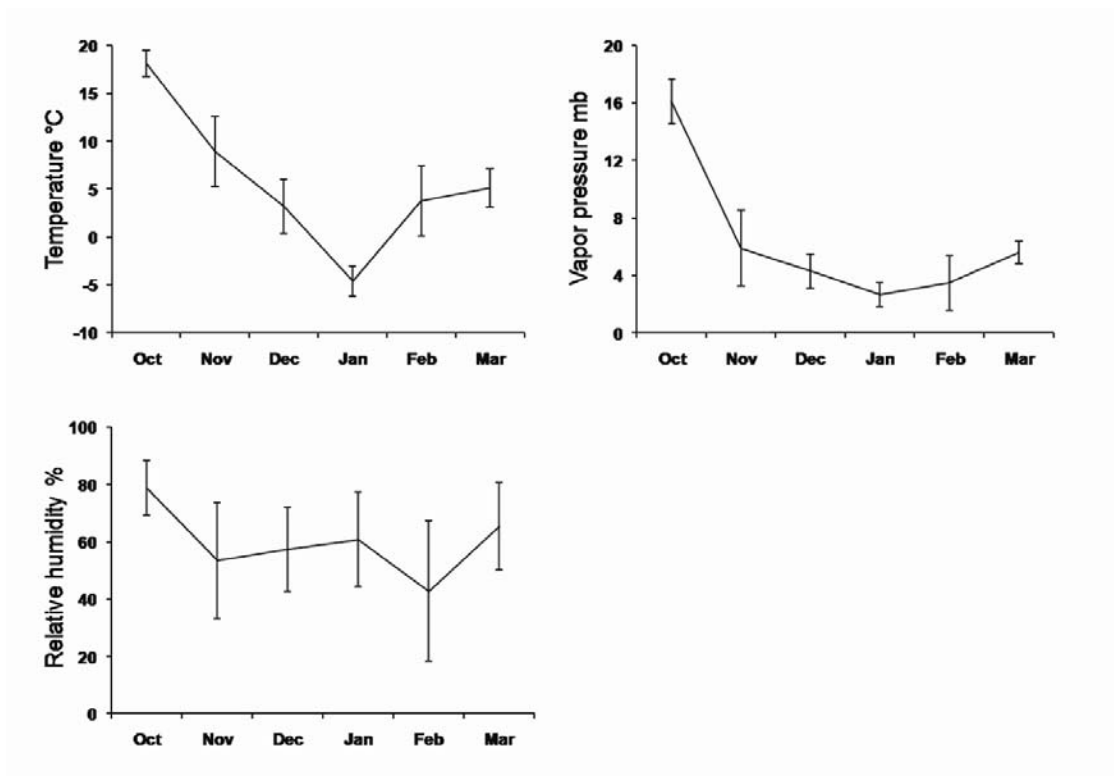


FIG. S4. Meteorological conditions (Temperature, $n = 48$; Vapor pressure, $n = 16$; Relative humidity, $n = 48$) at the RD site during air sampling period. Error bars represent standard deviation of means.

TABLE S1. Description of collected samples and corresponding meteorological conditions

<i>Land type</i>	<i>Date</i>	<i>Latitude, longitude</i>	<i>Elevation (m)</i>	<i>Sampling time (h)</i>	<i>AOP (mm)</i>	<i>Temp. (°C)</i>	<i>RH (%)</i>	<i>VP (mbar)</i>
RD	08/20~22 th /10 ^a	37°35' N 127°03' E	66	48	-	28.6 (2.3)	69.2 (11.0)	-
RD	10/10~12 th /10 ^b	37°35' N 127°03' E	66	48	-	18.1 (1.4)	78.8 (9.6)	16.1 (1.6)
RD	11/12~14 th /10 ^b	37°35' N 127°03' E	66	48	-	8.9 (3.6)	53.5 (20.2)	5.9 (2.6)
RD	12/21~23 th /10 ^b	37°35' N 127°03' E	66	48	-	3.2 (2.8)	57.4 (14.8)	4.3 (1.2)
RD	01/21~23 th /11 ^b	37°35' N 127°03' E	66	48	-	-4.6 (1.6)	60.8 (16.5)	2.7 (0.8)
RD	02/24~26 th /11 ^b	37°35' N 127°03' E	66	48	-	3.8 (3.7)	42.8 (24.6)	3.5 (1.9)
RD	03/27~29 th /11 ^b	37°35' N 127°03' E	66	48	-	5.1 (2.0)	65.5 (15.2)	5.6 (0.8)
RD-Rain	09/02 nd /10 ^a	37°35' N 127°03' E	66	-	37.0	24.1 (1.2)	83.0 (7.3)	-
FR	09/07~09 th /10 ^a	37°43' N 127°16' E	157	48	-	21.3 (1.9)	72.6 (13.3)	-
FR	11/03~05 th /10 ^b	37°43' N 127°16' E	157	48	-	6.6 (5.2)	81.6 (17.7)	-
IC	09/14~16 th /10 ^a	37°20' N 126°42' E	38	48	-	24.3 (2.9)	69.7 (19.1)	-
IC	11/16~18 th /10 ^b	37°20' N 126°42' E	38	48	-	6.0 (2.6)	75.0 (13.2)	-

Abbreviations: AOP, amount of precipitation; Temp, temperature; RH, relative humidity; VP, vapor pressure; RD, residential district;

RD-Rain, rainwater precipitated at the RD site; FR, forest; IC, industrial complex.

^aAir sampling for genotypic diversity of viruses.

^bAir sampling for viruses and bacteria enumeration.

For each type of meteorological conditions (Temp., $n = 48$; RH, $n = 48$; VP, $n = 16$), general mean values and s.d. (in parenthesis) are given.

TABLE S2. BLAST profile of four viromes compared with two public databases, *i.e.*, the CAMERA viral protein and CAMERA nr databases

	No. of sequences (%)							
	CAMERA_v				CAMERA_nr			
	BLASTx ($E < 10^{-3}$)				BLASTx ($E < 10^{-3}$)			
	Residential district (RD)	Forest (FR)	Industrial complex (IC)	Rainwater (RD-Rain)	Residential district (RD)	Forest (FR)	Industrial complex (IC)	Rainwater (RD-Rain)
Raw reads	24,721	23,316	20,578	18,829	24,721	23,316	20,578	18,829
Quality-filtered	18,427	20,903	14,754	15,914	18,427	20,903	14,754	15,914
Non-redundant	17,943 (100%)	20,691 (100%)	14,332 (100%)	15,776 (100%)	17,943 (100%)	20,691 (100%)	14,332 (100%)	15,776 (100%)
Unknown	10,590 (59.02%)	10,290 (49.73%)	11,463 (79.98%)	12,604 (79.89%)	10,683 (59.54%)	12,801 (61.87%)	12,491 (87.15%)	13,336 (84.53%)
Known	7,353 (40.98%)	10,401 (50.27%)	2,869 (20.02%)	3,172 (20.11%)	7,260 (40.46%)	7,890 (38.13%)	1,841 (12.85%)	2,440 (15.47%)
Virus	7,353 (40.98%)	10,401 (50.27%)	2,869 (20.02%)	3,172 (20.11%)	5,812 (32.39%)	7,074 (34.19%)	1,449 (10.11%)	1,508 (9.56%)
bacteria	-	-	-	-	1,407 (7.84%)	28 (0.14%)	236 (1.65%)	501 (3.18%)
Eukarya	-	-	-	-	40 (0.22%)	787 (3.80%)	143 (1.00%)	430 (2.73%)
Other	-	-	-	-	1 (0.01%)	1 (0.01%)	13 (0.09%)	1 (0.01%)

TABLE S3. Profiles of viral families from the four viromes compared with two public databases, *i.e.*, the CAMERA viral protein and CAMERA nr databases

		CAMERA_v				CAMERA_nr			
		BLASTx ($E < 10^{-3}$)				BLASTx ($E < 10^{-3}$)			
		Residential district	Forest	Industrial complex	Rainwater	Residential district	Forest	Industrial complex	Rainwater
		(RD)	(FR)	(IC)	(RD-Rain)	(RD)	(FR)	(IC)	(RD-Rain)
Virus sequences		7,353	10,401	2,869	3,172	5,812	7,074	1,449	1,508
dsDNA	Myo-	1	0	6	0	0	0	0	0
	Podo-	1	0	3	0	0	0	0	0
	Sipho-	3	0	199	3	1	0	21	0
	Unclassified Caudovirales	1	1	19	0	0	0	2	0
	Herpes-	1	0	0	0	0	0	0	0
	Polyoma-	0	0	11	0	0	0	0	0
	Pox-	0	0	18	0	0	0	16	0
	Unclassified-	42	11	23	7	28	2	0	0
	Unclassified dsDNA phages	2	9	7	7	1	2	0	0
	Unclassified dsDNA viruses	40	2	15	0	27	0	0	0
dsDNA viruses, no RNA stage	0	0	1	0	0	0	0	0	
Satellites	Satellites	3	16	9	2	3	8	5	3
ssDNA	Circo-	83	599	221	48	40	479	113	52
	Gemini-	144	952	598	652	82	317	357	98
	Ino-	2	0	1	0	0	0	0	0
	Micro-	90	79	433	29	21	21	78	2
	Nano-	42	32	18	23	39	24	7	18
	Unclassified-	6,721	8,335	888	1,834	5,492	5,949	604	1,282
	SsHADV-1	6,623	7,976	228	1,589	5,461	5,782	170	1,185
	Circovirus-like genomes	40	47	559	83	20	46	422	26
	ssDNA viruses	58	311	101	162	11	121	12	71
	ssRNA	Tombus-	1	0	0	0	1	0	0
Rhabdo-		0	0	4	0	0	0	4	0
Unclassified ssRNA positive-strand viruses		4	2	51	0	8	4	45	0
Unclassified	Unclassified-	214	374	367	574	97	270	197	53
	Unclassified phages	57	2	13	430	0	0	0	0
	Unclassified viruses	52	187	21	12	56	160	27	41
	Viruses	105	185	333	132	41	110	170	12

Abbreviations: SsHADV-1, *Sclerotinia sclerotiorum* hypovirulence-associated DNA virus 1.

TABLE S4. Nonanucleotide and stem sequences of the stem-loop structure found in airborne circoviruses and nanoviruses

Virome	Contig	Stem	Loop	Stem
Residential District (RD)	Contig35	AGTGTCCGGG	CT TAGTATTAC	CCCGGACACT
	Contig164	ACACGAGGAG	ATT TAGTATTAC	CTCCTCGTGT
Forest (FR)	Contig4	AGTGCCGG	TAC CAGTATTAC	CCGCCACT
	Contig26	GTGGCCAG	GAT TAGTATTAC	CTGGCCAC
	Contig133	CGAAGTCACGGG	TAGTATTAC	CCCGTGACTTCG
	Contig134	GGAAGTGACGG	C TAGTATTAC	CCGTCACTTCC
	Contig139	CACAAGTTGTCG	GCT CTATATTAC	CGACAACCTGTG
Industrial Complex (IC)	Contig119	GCGGGCTGG	CGT TAGTATTAC	CCAGCCCGC
	Contig126	AAGAGGTGG	CAT TAGTATTAC	CCACCTCTT
	Contig132	CAGAAGTGGCTG	CT CAGTATTAC	CAGCCACTTCTG
	Contig138	GGTGCCGG	AC CAGTATTAC	CCGGCACC
	Contig139	CCCAAAG	GT CAGTATTAC	CTTTTGGG