Virus Genes (2013) 46:441–446
DOI 10.1007/s11262-013-0878-2

Detection of African swine fever virus-like sequences in ponds in the Mississippi Delta through metagenomic sequencing

Xiu-Feng Wan · J. Lamar Barnett · Fred Cunningham · Si Chen · Guohua Yang · Shannon Nash · Li-Ping Long · Lorelei Ford · Sherry Blackmon · Yan Zhang · Larry Hanson · Qiang He

Received: 4 December 2012 / Accepted: 8 January 2013 / Published online: 22 January 2013

Abstract Metagenomic characterization of water virome was performed in four Mississippi catfish ponds. Although differing considerably from African swine fever virus (ASFV), 48 of 446,100 sequences from 12 samples were similar enough to indicate that they represent new members in the family Asfarviridae. At present, ASFV is the only member of Asfarviridae, and this study presents the first indication of a similar virus in North America. At this point, there is no indication that the identified virus(es) pose a threat to human or animal health, and further study is needed to characterize their potential risks to both public health and agricultural development.

Keywords African swine fever virus · Catfish pond · Freshwater · Metagenomics · Mississippi Delta · Asfarviridae

Freshwater bodies represent an important ecological interface between humans, mammals, waterfowl, reptiles, amphibians, fish, aquatic invertebrates, and insects, providing an optimal medium for the emergence of pathogens and subsequent intraspecies and interspecies transmission of infectious diseases. As a result, despite its economic and environmental importance, freshwater continues to be one of the most important media for transmission of infectious diseases worldwide. Examples of water as the medium for emerging and re-emerging waterborne pathogens include *Campylobacter* spp., *Cryptosporidium parvum*, *Legionella* spp., *Naegleria fowleri*, and *Burkholderia pseudomallei*, adenovirus, Norwalk and other caliciviruses, enteroviruses and reoviruses [2–5, 7, 8, 11, 15]. Particularly, the distribution and diversity of viruses remain largely under-studied in aquatic environments. Furthermore, the aquatic ecosystems provide a convenient concentration of pathogens from diverse host sources. For example, analyses of RNA viruses in a freshwater lake revealed large population diversity with the identification of novel viral sequences related to more than 30 viral families [1].

The Mississippi Delta is located in west Mississippi and bound by the Mississippi and Yazoo Rivers. This area represents one of the most important agricultural areas in the United States, producing more than 50 % of all farmed catfish in the country [14]. The Mississippi Delta is also known for its abundance in wildlife resources. As part of the Mississippi flyway, the region has sporadic high concentrations of diverse species of migratory birds. This region also has high concentrations of mammalian, reptilian, and amphibian species. With the intensity of catfish farming, the large numbers of catfish ponds in the Mississippi Delta frequented by migratory bird populations serve as an important ecological intersection between humans, catfish, migratory waterfowls, and other wild animal populations. Given the potential importance of catfish ponds in the distribution and transmission of microbial populations, we used a metagenomic approach to characterize the microbial profiles that occur in this region and to identify any potential microbial threats to public and

X.-F. Wan (✉) · J. L. Barnett · G. Yang · S. Nash · L.-P. Long · L. Ford · S. Blackmon · L. Hanson
Department of Basic Sciences, College of Veterinary Medicine, Mississippi State University, Mississippi state, MS, USA
e-mail: wanhenry@yahoo.com; wan@cvm.msstate.edu

F. Cunningham
USDA/APHIS/WS, National Wildlife Research Center, Mississippi Field Station, Mississippi state, MS, USA

S. Chen · Y. Zhang · Q. He
Department of Civil and Environmental Engineering, The University of Tennessee, Knoxville, TN, USA

Table 1 Sample collection at the Mississippi Delta

Sample index	Date	Pond	454 reads	Comments
S1	Jan 21, 2010	1	87,962	Inactive pond
S2	Jan 21, 2010	2	15,585	Inactive pond
S3	Jan 21, 2010	3	41,328	Active pond
S4	Jan 21, 2010	4	21,228	Active pond
S5	Feb 25, 2010	3	41,908	Active pond
S6	Feb 25, 2010	4	32,398	Active pond
S7	Feb 25, 2010	1	63,198	Inactive pond
S8	Feb 25, 2010	2	43,137	Inactive pond
S9	Apr 9, 2010	1	24,375	Inactive pond
S10	Apr 9, 2010	2	23,940	Inactive pond
S11	Apr 9, 2010	3	13,381	Active pond
S12	Apr 9, 2010	4	37,660	Active pond

animal health in catfish ponds in the Mississippi Delta region.

We obtained 12 samples from four ponds from two sites approximately 30 miles apart at Stoneville, MS. Two ponds were from an active fish farm with mowed banks and heavy traffic nearby, whereas the other two were from an inactive/abandoned farm. A total of 80 L water samples were collected from each pond in January, February, and April of 2010, respectively (Table 1). Samples were concentrated by tangential flow ultrafiltration followed by secondary concentration with Centricon Plus-70 centrifugal filtration units (Millipore; 30,000 Daltons-cutoff) as previously described [11, 12]. Total nucleic acids were extracted from each sample and then subjected to 454 pyrosequencing. A total of over 446,000 sequence reads were obtained, each measuring 15,585–87,962 reads (Table 1). The lengths of these sequences varied from 70 to 332 nucleotides. Of these reads, 52 % were bacteria, 6.9 % were viruses, and 27.9 % did not match any known sequences. Among the sequences identified as viruses, 86.97 % belong to Caudovirales, 4.11 % to Phycodnaviridae, 2.38 % to unclassified double stranded DNA phages, and 1.4 % to unclassified dsDNA viruses.

Further analyses showed 48 of the sequences are similar to African swine fever virus (ASFV) of the family Asfarviridae. These 48 sequences showed homology to 23 genes of ASFV, including capsid, helicase, and DNA polymerase (Table 2). Sequence and phylogenetic analyses demonstrate these sequences are most closely related to the viral sequences from Asfarviridae family but are highly divergent from those reported ASFV sequences or other ASFV-like sequences. We propose these novel sequences could be derived from new member(s) in the Asfarviridae family, and this represents the first potential Asfarviridae identified in North America.

The amino acid sequences translated from these 48 matches are highly divergent from the known ASFV sequences. The BLASTX e-values varied from 0.11 to 3.00E−025, and BLASTX searches protein databases using a translated nucleotide query (Table 2). The translated protein sequences shared high identities, varying from 28.1 to 64.9 %, with 23 genes in ASFV. Seven reads align with five regions of capsid protein B646L without overlaps between these samples (Table 2). Through analysis using the Pfam database, these capsid-like genes were shown to have Capsid_NCLDV (Large eukaryotic DNA virus major capsid protein) domain with E-value 0.00012 [13]. The sequence alignment showed this domain was expressed across capsid genes from different families (Fig. 1a). Similarly, nine sequences were similar to two individual helicase genes: three reads to B962L and six to D1133L (Table 2; Fig. 1b). Using the Pfam database, the SNF2 family N-terminal domain was identified in these sequences with E-value 2.7E−7. There was a single read similar to alpha-like DNA polymerase (G1211R), with BLASTX E-value 3.00E−018 and 39.3 % protein sequence identity. Phylogenetic analysis using Maximum Parsimony method implemented in PAUP*4.0 demonstrated that the newly identified sequences coding capsid and helicase genes are genetically most closely related to the ASFV isolates in the Asfarviridae family (Fig. 2).

The GC content for ASFV is about 38.42 %, and the GC contents of capsid gene and helicase gene in ASFV are 42.33 and 46.47 %, respectively. There are large variations in the GC contents of the ASFV-like sequences we identified. For example, the seven identified capsid-like genes have GC contents varying from 31.99 to 50.96 %, with an average of 43.00 %; the nine identified helicase-like genes have GC contents varying from 27.72 to 56.65 %, with an average of 40.40 %.

Two pairs of primers (capsid: SFC2-F: AGGTTCA TCACTATTGACCTC and SFC2-R: GGTCGTTATTG ACTGTGACTC and helicase: SFH1-F: CTTTCAGTTC TATGGCTATCGTC and SFH1-R: TCATTATTGCTAG ACCACTCAG) were designed for the capsid and helicase-like sequences, and their presence was confirmed by PCR. PCRs were performed in the first four samples (samples 1 through 4) (not done for the other samples due to lack of enough DNA), which identified both capsid-like and helicase-like genes in sample 1 and shared positive 454 reads. PCR identified helicase-like sequence and capsid-like sequences in sample 4, which were not found to be positive for this virus detected in metagenomic analyses. Thus, our results suggest that ASFV-like segments were present in both active and inactive catfish ponds. The PCR products were then sequenced, and the resulting sequences were identical to those identified from 454 sequencing. The failure to identify these genes in other samples in 454 pyrosequencing suggests that additional deeper sequencing will help recover more reads.

Table 2 Summary of the ASFV-like sequence identified from the catfish ponds in Mississippi Delta

Sample	Hit	Gene	Start	Stop	E-value	Identity (%)	Function
S1	1	D1133L	126362	126285	3.00E−005	50.0	Helicase superfamily II
	2	S273R	130496	130606	4.00E−003	35.1	Polyprotein processing protease
	3	S273R	130496	130582	3.00E−004	37.9	Polyprotein processing protease
	4	B602L	84255	84016	3.00E−014	32.5	Capsid chaperone
	5	A276R	32381	32536	1.00E−005	32.7	Multigene family 360
S7	1	D205R	127963	128100	8.00E−010	34.8	RNA polymerase subunit 5
	2	G1211R	97022	97273	3.00E−018	39.3	Alpha-like DNA polymerase
	3	B646L	87500	87366	3.00E−009	37.8	Major capsid protein
	4	EP1242L	50427	50299	2.00E−015	51.2	RNA polymerase subunit 2
	5	EP1242L	51435	51244	5.00E−022	46.9	RNA polymerase subunit 2
	6	D1133L	123782	123693	3.00E−002	36.7	Helicase superfamily II
	7	B646L	88595	88530	2.00E−003	45.5	Major capsid protein
	8	CP530R	108895	109065	8.00E−003	28.1	pp62 Polyprotein precursor of structural proteins p35 and p15
	9	P1192R	132729	132863	3.00E−010	35.6	Topoisomerase II
	10	D1133L	125033	124869	7.00E−023	47.3	Helicase superfamily II
	11	B354L	83485	83577	0.34	32.3	ATP or GTP binding motif
	12	B646L	88733	88626	3.00E−008	50.0	Major capsid protein
	13	M1249L	61272	61036	1.00E−008	30.4	Putative
	14	B354L	83277	83107	0.11	28.1	ATP or GTP binding motif
	15	G1340L	95145	94981	1.00E−005	30.9	Vaccinia A8L-like transcription factor
	16	S273R	130079	130258	5.00E−016	41.7	Polyprotein processing protease
	17	CP2475L	105195	105091	9.00E−016	48.6	pp20 Polyprotein precursor of structural proteins p150, p37, p14 and p34
	18	CP2475L	103872	103729	7.00E−018	50.0	Two RGD cell attachment motifs
	19	NP1450L	114788	114456	3.00E−023	36.0	RNA polymerase subunit 1
	20	B962L	77552	77466	1.00E−013	51.7	Helicase superfamily II
	21	B962L	77615	77547	9.00E−009	47.8	Helicase superfamily II
S10	1	CP2475L	105213	105091	3.00E−007	43.9	pp20 Polyprotein precursor of structural proteins
	2	NP868R	120362	120643	2.00E−016	34.0	Guanylytransferase
	3	NP1450L	112667	112575	4.00E−004	48.4	RNA polymerase subunit 1
	4	D250R	120847	120918	1.00E−004	45.8	Nudix hydrolase
	5	S273R	130559	130621	1.00E−002	57.1	Cell attachment sequence RGD
	6	EP1242L	50379	50299	8.00E−005	48.1	RNA polymerase subunit 2
S12	1	B646L	87764	87618	4.00E−017	57.1	Major capsid protein
	2	D1133L	126683	126573	6.00E−017	64.9	Helicase superfamily II
	3	B646L	88595	88419	3.00E−025	52.5	Major capsid protein
	4	G1340L	95997	95851	5.00E−007	38.8	Vaccinia A8L-like transcription factor
	5	B646L	87503	87387	2.00E−018	61.5	Major capsid protein
	6	B646L	96045	95905	7.00E−006	34.0	Major capsid protein
	7	M1249L	59355	59215	1.00E−004	31.9	Putative gene
	8	B602L	84453	84307	5.00E−010	38.8	Capsid chaperone
	9	D1133L	125843	125733	1.00E−009	48.6	Helicase superfamily II
	10	D1133L	126134	126000	2.00E−022	46.7	Helicase superfamily II
	11	C717R	65797	66039	3.00E−010	29.6	Putative
	12	E66L	152706	152635	6.00E−005	41.7	Putative
	13	Q706L	141681	141445	6.00E−019	36.7	Helicase superfamily II

Table 2 continued

Sample	Hit	Gene	Start	Stop	E-value	Identity (%)	Function
	14	B962L	77360	77157	9.00E−015	39.7	Helicase superfamily II
	15	B385R	86386	86559	4.00E−009	34.5	Putative product contains Zn finger C2H2
	16	C475L	70011	69814	3.00E−006	30.3	Poly A polymerase large subunit

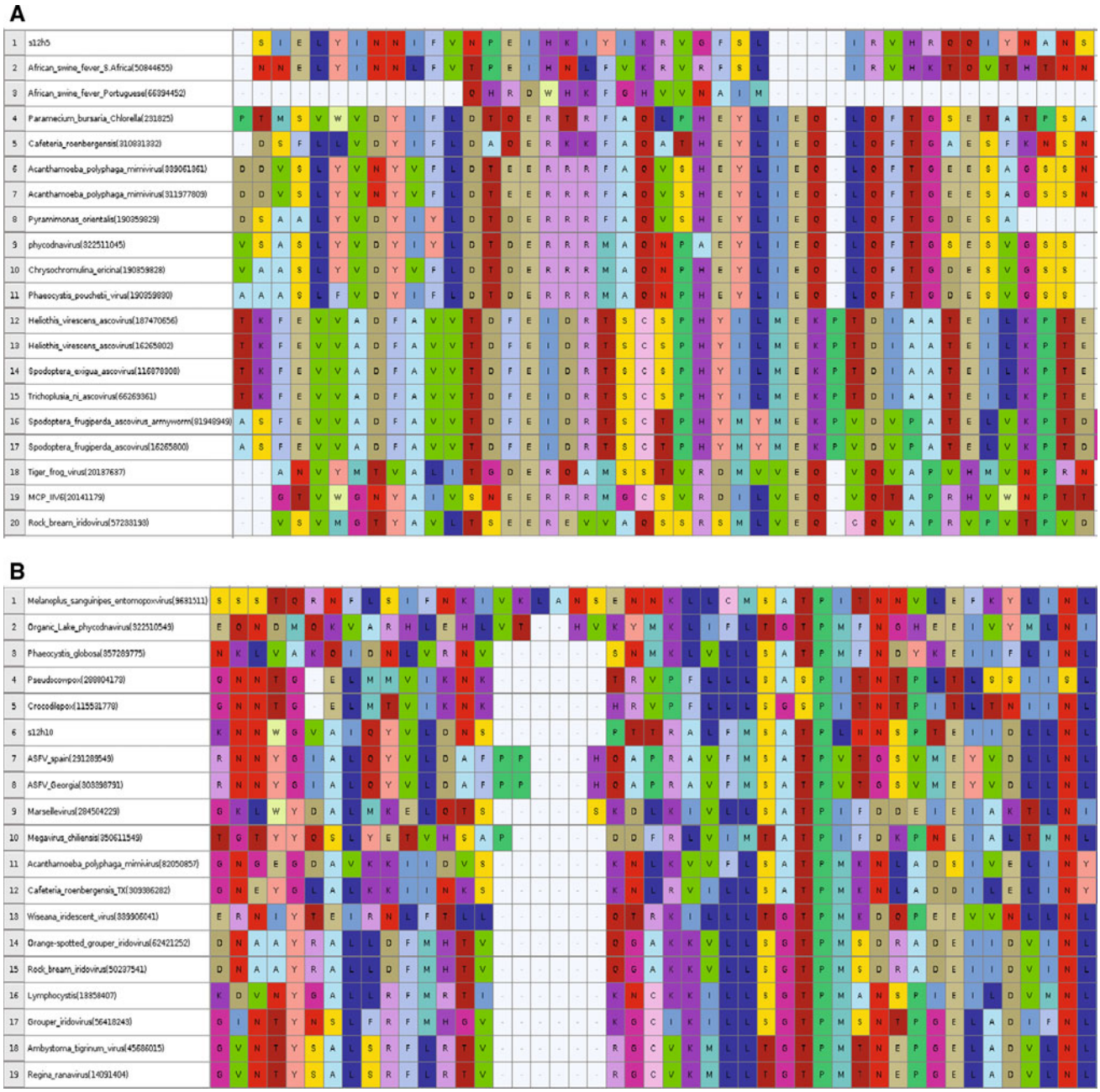

Fig. 1 Alignment of capsid-like (**a**) and helicase-like (**b**) protein sequences. The *numbers* in the *parenthesis* after each species is the GenBank accession number for the capsid or helicase used in this species (Color figure online)

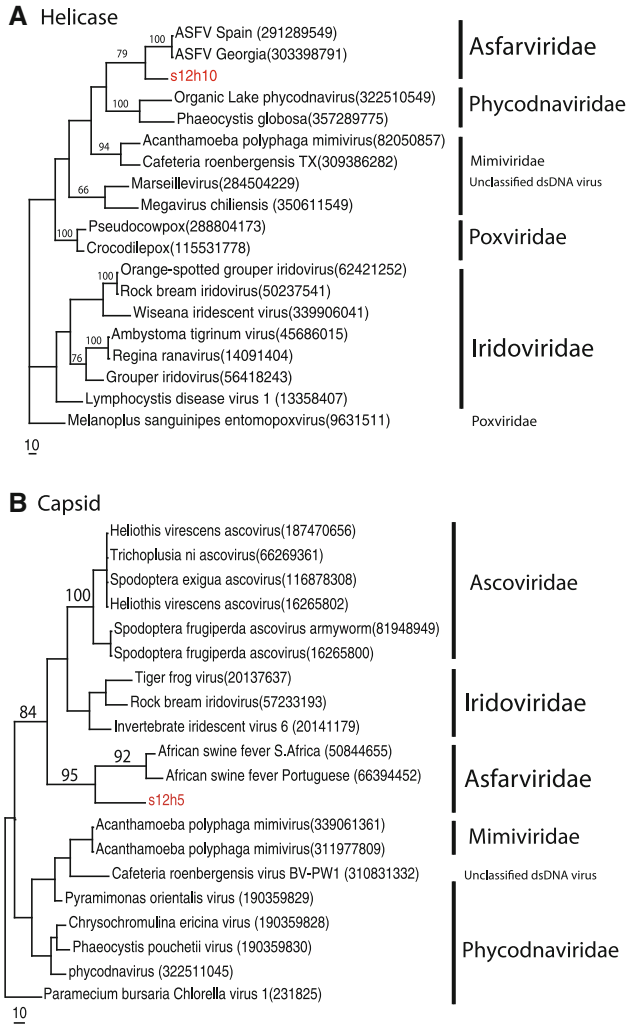

Fig. 2 Phylogenic analysis of the capsid-like (**a**) and helicase-like (**b**) sequences. The *numbers in the parentheses* denote the GenBank accession numbers. The sequences (s12h5 and s12h10) were marked in *red* (Table 1) (Color figure online)

African swine fever virus is the only known virus in the family of Asfarviridae. ASFV is the causative agent of hemorrhagic viremia in domestic swine. This pathogen is endemic to Africa and naturally occurs in wild swine species without causing disease. The natural life cycle involves transmission by soft ticks. It is the only known DNA virus that is transmitted by arthropods. Outside of Africa, outbreaks of ASFV have been documented in Europe and Central/South America. These outbreaks were linked to the movement of domestic pigs. After the outbreaks in America, ASFV was eliminated by depopulation. Our metagenomic results suggest that there is more than one member in the family Asfarviridae and that at least one member of Asfarviridae is present in North America. To better characterize this new virus, PCRs were performed, which confirmed the positive amplification of ASFV-like capsid and helicase gene sequences. However, PCR amplification targeting the regions between these two genes, which neighbor each other in ASFV (spanning 2.2 kb), was not successful (data not shown). Thus, the gene order in this potential new virus is not necessarily the same as that in ASFV.

Interestingly, a recent study has also identified ASFV-like sequences from human serum and sewage [6]. These reported sequences shared sequence identities from 27 to 64 % with ASFV. Comparison of the 48 sequences from our study with those recovered by Loh et al. [6] indicate that the homologous regions of these sequences are located in different regions of the ASFV genome, and further study is required to understand the natural host potential and pathogenicity of this virus. Another study reported ASFV-like sequencers were identified in human sera from patients with dengue-like symptoms but testing negative for dengue virus [16]. In addition, metagenomic data demonstrated that ASFV-like sequences were present in marine environments [9, 10]. These studies support our findings in this study that it is more likely that new members are present in the family Asfarviridae. Mississippi has diverse vertebrate wildlife including a large population of feral pigs [2], but since emerging ASFV is a tick-borne virus, an invertebrate host is also likely. At this point, there is no indication that the identified virus poses a threat to human or animal health.

Acknowledgments We thank Mary Duke and Dr. Brian Scheffler at USDA-ARS, Mid-South Area Genomics Laboratory, Stoneville, MS for genomic sequencing. This Project was supported by USDA APHIS NWRC Award 1174280807CA 12030306 and NIH RC1AI086830 to XFW.

References

1. A. Djikeng, R. Kuzmickas, N.G. Anderson, D.J. Spiro, Metagenomic analysis of RNA viruses in a fresh water lake. PLoS ONE **4**, e7264 (2009)
2. T.J. Inglis, S.C. Garrow, M. Henderson, A. Clair, J. Sampson, L. O'Reilly, B. Cameron, *Burkholderia pseudomallei* traced to water treatment plant in Australia. Emerg. Infect. Dis. **6**, 56–59 (2000)
3. J.E. Kaplan, R.A. Goodman, L.B. Schonberger, E.C. Lippy, G.W. Gary, Gastroenteritis due to Norwalk virus: an outbreak associated with a municipal water system. J. Infect. Dis. **146**, 190–197 (1982)
4. H. Leclerc, L. Schwartzbrod, E. Dei-Cas, Microbial agents associated with waterborne diseases. Crit. Rev. Microbiol. **28**, 371–409 (2002)
5. W.J. Lodder, H.H. van den Berg, S.A. Rutjes, A.M. de Roda Husman, Presence of enteric viruses in source waters for drinking water production in the Netherlands. Appl. Environ. Microbiol. **76**, 5965–5971 (2010)
6. J. Loh, G. Zhao, R.M. Presti, L.R. Holtz, S.R. Finkbeiner, L. Droit, Z. Villasana, C. Todd, J.M. Pipas, B. Calgua, R. Girones, D. Wang, H.W. Virgin, Detection of novel sequences related to African swine fever virus in human serum and sewage. J. Virol. **83**, 13019–13025 (2009)

7. W.R. MacKenzie, J.J. Kazmierczak, J.P. Davis, An outbreak of cryptosporidiosis associated with a resort swimming pool. Epidemiol. Infect. **115**, 545–553 (1995)
8. K.D. Mena, C.P. Gerba, Waterborne adenovirus. Rev. Environ. Contam. Toxicol. **198**, 133–167 (2009)
9. A. Monier, J.M. Claverie, H. Ogata, Taxonomic distribution of large DNA viruses in the sea. Genome Biol. **9**, R106 (2008)
10. H. Ogata, K. Toyoda, Y. Tomaru, N. Nakayama, Y. Shirai, J.M. Claverie, K. Nagasaki, Remarkable sequence similarity between the dinoflagellate-infecting marine girus and the terrestrial pathogen African swine fever virus. Virol. J. **6**, 178 (2009)
11. P. Payment, M. Trudel, S.A. Sattar, V.S. Springthorpe, T.P. Subrahmanyan, B.E. Gregory, A.H. Vajdic, P. Blaskovic, I.J. Guglielmi, O. Kudrewko, Virological examination of drinking water: a Canadian collaborative study. Can. J. Microbiol. **30**, 105–112 (1984)
12. A.L. Polaczyk, J. Narayanan, T.L. Cromeans, D. Hahn, J.M. Roberts, J.E. Amburgey, V.R. Hill, Ultrafiltration-based techniques for rapid and simultaneous concentration of multiple microbe classes from 100-L tap water samples. J. Microbiol. Methods **73**, 92–99 (2008)
13. M. Punta, P.C. Coggill, R.Y. Eberhardt, J. Mistry, J. Tate, C. Boursnell, N. Pang, K. Forslund, G. Ceric, J. Clements, A. Heger, L. Holm, E.L. Sonnhammer, S.R. Eddy, A. Bateman, R.D. Finn, The Pfam protein families database. Nucleic Acids Res. **40**, D290–D301 (2012)
14. C.S. Tucker, The ecology of channel catfish culture ponds in Northwest Mississippi. Rev. Fish. Sci. **4**, 1–55 (1996)
15. R.L. Vogt, H.E. Sours, T. Barrett, R.A. Feldman, R.J. Dickinson, L. Witherell, Campylobacter enteritis associated with contaminated water. Ann. Intern. Med. **96**, 292–296 (1982)
16. N.L. Yozwiak, P. Skewes-Cox, M.D. Stenglein, A. Balmaseda, E. Harris, J.L. DeRisi, Virus identification in unknown tropical febrile illness cases using deep sequencing. PLoS Negl. Trop. Dis. **6**, e1485 (2012)