environmental microbiology

Environmental Microbiology (2016) **18**(4), 1162–1175 doi:10.1111/1462-2920.13127

Analysis of dsDNA and RNA viromes in methanogenic digesters reveals novel viral genetic diversity

Magdalena Calusinska,[1†] Martyna Marynowska,[1,2†] Xavier Goux,[1] Esther Lentzen[3] and Philippe Delfosse[1*]

[1]*Luxembourg Institute of Science and Technology, Environmental Research and Innovation (ERIN), 41 rue du Brill, L-4422, Belvaux, Luxembourg.*

[2]*Intercollegiate Faculty of Biotechnology, University of Gdansk and Medical University of Gdansk, ul. Kladki 24, 80-822, Gdansk, Poland.*

[3]*Luxembourg Institute of Science and Technology, Materials Research and Technology (MRT), 41 rue du Brill, L-4422, Belvaux, Luxembourg.*

Summary

Although viruses are not the key players of the anaerobic digestion process, they may affect the dynamics of bacterial and archaeal populations involved in biogas production. Until now viruses have received very little attention in this specific habitat; therefore, as a first step towards their characterization, we optimized a virus filtration protocol from anaerobic sludge. Afterwards, to assess dsDNA and RNA viral diversity in sludge samples from nine different reactors fed either with waste water, agricultural residues or solid municipal waste plus agro-food residues, we performed metagenomic analyses. As a result we showed that, while the dsDNA viromes (21 assigned families in total) were dominated by dsDNA phages of the order *Caudovirales*, RNA viruses (14 assigned families in total) were less diverse and were for the main part plant-infecting viruses. Interestingly, less than 2% of annotated contigs were assigned as putative human and animal pathogens. Our study greatly extends the existing view of viral genetic diversity in methanogenic reactors and shows that these viral assemblages are distinct not only among the reactor types but also from nearly 30 other environments already studied, including the human gut, fermented food, deep sea sediments and other aquatic habitats.

Received 19 December, 2014; revised 17 August, 2015; accepted 6 November, 2015. *For correspondence. E-mail philippe.delfosse@list.lu; Tel. 00352 470261 442; Fax 00352 275885. †Equal contribution.

Introduction

Although bacteria and archaea are the key players of the anaerobic digestion (AD) process occurring in methanogenic reactors producing biogas, the dynamics of these communities may be affected by numerous environmental and biological factors (Chen *et al.*, 2008). Bacteriophage predation can regulate microbial abundance and diversity in full-scale biogas units (Shapiro *et al.*, 2010), according to the proposed 'phage kills the winner' strategy (Thingstad and Lingen, 1997). Even though the predator–prey paradigm is still underdeveloped in microbial ecology, research dedicated to aquatic habitats estimated that viral activity (e.g. specific lytic infection and/or prophage induction and activation of the viral DNA replication) destroys daily around 20% of microbial cells worldwide, thus having important implications to microbial food chains, carbon turnover and the global carbon cycle (Suttle, 2007; Maurice *et al.*, 2013).

Viruses, being the most abundant biological units on the planet, reaching 10^{31} particles globally (Hendrix, 2002), have been detected at concentrations 10–1000 times higher in waste water treatment plants (WWTP), including AD reactors, than in any other aquatic habitat (Wommack and Colwell, 2000; Wu and Liu, 2009). Nevertheless, the AD viral population (virome) has received very little attention compared with the AD microbial population (microbiome). Even though bacteriophage lysis was shown to regulate microbial abundance and diversity in full-scale AD reactors treating industrial waste water (Shapiro *et al.*, 2010), and phages were shown to propagate in methanogenic reactors at a rate in terms of virus-like particles (VLPs) of at least 5.2×10^7 VLPs day^{-1} l^{-1} (Park *et al.*, 2007), to date there are only a few studies characterizing the viral genomic content (mainly dsDNA phages) in these environments (Wu and Liu, 2009; Tamaki *et al.*, 2012; Bibby and Peccia, 2013a). In the case of RNA viruses, no study has ever been conducted in a methanogenic digester.

Here, in addition to the morphological perspective, we present a detailed metagenomic characterization of dsDNA and RNA viruses present in AD reactors and reveal their huge genetic diversity in nine different sludge samples collected from WWTP, farm digesters and biogas units treating a mixture of municipal solid waste and agro-food residues. We believe that using an optimized VLP

filtration protocol that excluded the multiple displacement amplification (MDA) commonly used in viral metagenomics (MDA is known to enhance chimera formation and to create random over-amplification; Marine *et al*., 2014), together with a combination of both sequence-independent and sequence-dependent data analyses, allowed us to establish a detailed and fair picture of viral assemblages in AD reactors.

Results and discussion

Optimization of the VLPs filtration protocol from sludge

Sludge samples were collected from nine anaerobic reactors treating, respectively, (i) waste water, (ii) agricultural residues, and (iii) agricultural/agro-food residues and municipal solid waste (Table 1). To evaluate the impact of the pre-filtration and the tangential flow filtration (TFF) protocols on the efficiency of VLPs recovery, a representative sludge sample was autoclaved (to keep the sample texture close to the original but to eliminate viruses) and spiked in with a ΦX174 phage suspension. The phage ΦX174, spiked at a known amount, was used to evaluate the recovery rate of VLPs at the different steps of the virus filtration protocol (Table S1). According to the number of plaque-forming units (pfu ml^{-1}) calculated for each fraction, the amount of the phage remained in the same order of magnitude after the pre-filtration steps. However, it decreased by onefold following the TFF, mounted with a Sartocon Slice 200 Hydrosart cassette of 0.22 μm, and ultracentrifugation steps. Longer ultracentrifugation time did not result in a higher phage recovery rate (results not shown). Although in general the VLPs filtration protocol performed well, the two last steps could introduce a quantitative bias, especially important if absolute quantification of viruses in a sample is considered (not the case in our study).

To further evaluate the differential detection of known viral families and the potential loss of some of the largest viruses that may exceed the size of 0.22 μm (Abrahao *et al*., 2014), another TFF with a cassette of 0.45 μm size was tested using metagenomics on two selected samples (P6_2 and C7_2). In our case, no additional families were detected for dsDNA viromes with the cassette of 0.45 μm size in comparison to 0.22 μm, and the data for the two samples analysed with the two cassette types correlated well (Fig. S1). Moreover, the estimated total species richness, based on contiguous sequences (contigs) spectra, was independent of the cassette pore size used (Table 2). For RNA viromes no striking differences at the family level were observed for either of the two cassette types used. Surprisingly, the estimated viral richness was slightly higher for the 0.22 μm cassette.

A huge diversity of viral morphologies is observed in AD reactors by means of transmission electron microscopy (TEM)

Using TEM results, the majority of purified VLPs resembled well-known viral morphologies, with filamentous viruses, including flexuous filaments similar to *Inovirus*-like or *Tymovirales* VLPs and also rigid rod-shaped particles (e.g. *Tobamovirus*-like) being intermixed with tailed and tailless icosahedral VLPs of different size ranges. Ovoid or bacilliform morphology was also observed (Fig. 1). Head-tailed and tailless icosahedral viruses infecting both bacteria and archaea have been previously described in methanogenic environments (Pietila *et al*., 2014). *Siphoviridae*-like head-tailed viruses have already been reported for methanogenic digesters (Chien *et al*., 2013), and their presence is consistent with the dsDNA viral genetic diversity discussed below and in a previous metagenomic study (Tamaki *et al*., 2012).

AD reactors constitute a reservoir of novel viral genetic diversity

The initial sequence-independent comparative analysis of metagenomic reads proposed in this study was similar to the concept developed in crAss, where cross-metagenome assembly of reads enables the comparison of entities (e.g. contigs) that are shared between samples (Dutilh *et al*., 2012). However, the use of the CLC Genomic Workbench instead of crAss allowed us to perform the co-assembly (separately for dsDNA and RNA viromes) and to determine a normalized (per metagenome size and contig length) degree of relatedness (here referred to contigs spectra), within the same application (not the case for crAss that requires a separate assembly software to be used). In continuation, these relative contigs abundance spectra expressed as DNA-reads per kilobase of transcript (gene) per million reads mapped (DNA-RPKM; here equalled to the number of reads mapped to the contig and normalized by the contig length and per million mappable reads), together with a contig affiliation (sequence-dependent approach), gave a view on a normalized composition profile of dsDNA and RNA viromes for each sludge sample.

In this study over 42 million (M) and 25 M raw reads, generated, respectively, for 11 and 10 metagenomic libraries prepared for dsDNA and RNA viromes (Table S2), were quality trimmed and co-assembled, taking advantage of the fact that virus-like sequences rare in one sample can be abundant in another, thus leading to improved contigs (Minot *et al*., 2013). When the reads were mapped back to the assembled contigs, only 48.5% could be aligned for dsDNA viromes, and the collector's curve analysis showed that the detection of the potential viruses-like sequences was not completely saturated

Table 1. Origin of the AD sludge samples and their characteristics.

	AD unit	AD reactor size (m^3)	Main feeding substrate	Sample name	Collection date	pH	Total inorganic carbon[a]	Total NH_4-N[b]	Total solid[c]	Volatile solid[d]
Waste water treatment plant	Schifflange LU	2500	Sewage sludge from WWTP treatment	S3	November 2012	7.29	1.20	1.82	3.59	54.93
	Pétange LU	1350		P6, P6_2	December 2013	7.17	1.45	0	2.40	52.42
	Bettembourg LU	1500		BT5	December 2013	7.07	1.30	0.42	3.00	49.16
Agricultural wastes	Unit 1	450	Maize and grass silage, cow manure and slurry, bread	B10	December 2013	8.06	4.90	4.20	7.47	73.93
	Unit 2	500	Maize silage, cow manure	C7, C7_2	December 2013	8.15	4.70	4.20	10.62	72.95
	Unit 3	1270	Cow manure and slurry, grass, maize and cereal silage	BKA7	December 2013	7.69	2.30	0.83	7.53	75.62
Agricultural/agro-food residues and municipal solid waste	Unit 4	1900	Yard green waste (leaves, trees), organic wastes from supermarkets and canteens	MK18	January 2014	7.83	2.50	3.22	22.01	54.78
	Unit 5	1100	Manure, cereal waste, grass silage, organic wastes from supermarkets and canteens	MbioA6	December 2013	7.65	3.70	2.80	6.43	65.67
	Unit 6	4000	Cow manure and slurry, cereal waste, wastes from food industries, maize silage, grass	F4	January 2013	7.84	4.00	1.26	8.24	60.71

Samples with a 'name_2' were prepared with a 0.45 μm size filter.
a. m^3 of CO_2/m^3 of sludge.
b. kg NH4-N/m^3 of sludge.
c. % of total solid/ fresh sludge (w/w).
d. % of volatile solid/ % of total solid (w/w).

Table 2. Estimated species richness for DNA and RNA viromes calculated with CatchAll.

		dsDNA viromes			RNA viromes			
	Sample	Best model	Estimated total species ±SE	Best discounted model	Estimated total species ±SD	Best model	Estimated total species ±SE	Best discounted model	Estimated total species ±SD
Waste water treatment plant	S3	FourMixedExp	12 784 ± 165	ThreeMixedExp	8364 ± 108	SingleExp	142.6 ± 1.3	–	
	P6	FourMixedExp	13 052 ± 189	–		TwoMixedExp	111.7 ± 2.3	SingleExp	83.3 ± 1.6
	P6_2	FourMixedExp	12 775 ± 56	ThreeMixedExp	7846 ± 35	SingleExp	72.8 ± 2.4	–	
	BT5	ThreeMixedExp	8345 ± 47	TwoMixedExp	2186 ± 12	SingleExp	92.2 ± 4	–	
Agricultural wastes	B10	FourMixedExp	18 257 ± 3328	–		TwoMixedExp	47.4 ± 5.1	SingleExp	26 ± 1.5
	C7	FourMixedExp	15 685 ± 203	ThreeMixedExp	7445 ± 96	SingleExp	22.8 ± 2	–	
	C7_2	FourMixedExp	17 559 ± 361	ThreeMixedExp	8490 ± 174	SingleExp	17.6 ± 1.4	–	
	BKA7	FourMixedExp	14 076 ± 277	–		SingleExp	28 ± 2	–	
Agricultural/agro-food residues and municipal solid waste	MK18	FourMixedExp	14 802 ± 315	ThreeMixedExp	9318 ± 198	ND			
	MbioA6	ThreeMixedExp	12 631 ± 163	–		SingleExp	61.4 ± 1.2	–	
	F4	FourMixedExp	15 922 ± 336	ThreeMixedExp	12 068 ± 254	SingleExp	46.5 ± 0.8	–	

ND, not done. Where no richness value is indicated, the software failed to calculate the species richness for the best discounted model. Samples with a 'name_2' were prepared with a 0.45 µm size filter.

(Fig. S2). In contrast, nearly 82% of reads could map back to the assembled RNA contigs, and the calculated rarefaction curves confirmed that most of the existing viral RNA genetic diversity in AD reactors of the three main types was characterized in this study (Fig. S2).

For dsDNA viromes a co-assembly of reads yielded over 92 K contigs, while only a third (27.9 K) was taxonomically assigned of viral origin (Tables S3 and S4). To verify a possible microbial genomic DNA (gDNA) contamination versus potentially novel massive diversity of dsDNA viruses in AD reactors, the predicted protein features for virus-assigned and unassigned dsDNA contigs were further functionally annotated (Fig. 2A and B). As a result, for contigs assigned of viral origin, 56.3% of the predicted proteins could have been functionally annotated, and their functional profile was different from the non-virus assigned contigs (Pearson correlation coefficient, r of 0.375). Using the subsystems approach (Overbeek *et al*., 2005), 30.95% of annotated proteins fell into the category 'phages, prophages, transposable elements, plasmids', and 14.75% were classified into 'clustering-based subsystems' (unknown function), followed by 'DNA metabolism' (10.46%). In the case of the remaining contigs (64.8 K) that could not be taxonomically assigned of viral origin, only 26.5% of their predicted proteins could have been functionally annotated (Fig. 2A). Their distribution into subsystems categories was similar to a typical microbial metagenome from an AD digester (r of 0.994; Magdalena Calusinska unpubl. data), with three dominant subcategories, including 'clustering-based subsystems' (16.22%), 'protein metabolism' (11.27%) and 'carbohydrates' (11.25%), pointing to bacterial gDNA contamination. Interestingly, the remaining predicted unknown protein features for virus-unassigned contigs that could not be annotated point towards a novel genetic diversity potentially of viral origin. Indeed, out of the estimated 3.9 M protein clusters for the global virome (Ignacio-Espinoza *et al*., 2013), only around 5000 viral genomes were sequenced and annotated at the time the study was conducted. Moreover, as the majority of the determined open reading frames (ORFs) have no close relatives available in the public databases (Roux *et al*., 2014a), the analysis of AD viromes provides strong evidence of novel viral genetic diversity in this environment.

Four DNA contigs were ≥ 40 kb in size, with the longest one being 48 kb. Two between the largest DNA contigs (contig_ 25076 encodes 61 genes and contig_23253 encodes 58 genes), representing unclassified *Siphoviridae* phages, were putatively complete linear phage genomes, and no partially sequenced genes were detected at both ends (Fig. 3). Indeed, previous studies investigating viral assemblages in methanogenic reactors reported the size of some viral DNA genomes to be between 30 and 80 kb (Wu and Liu, 2009), which is in the

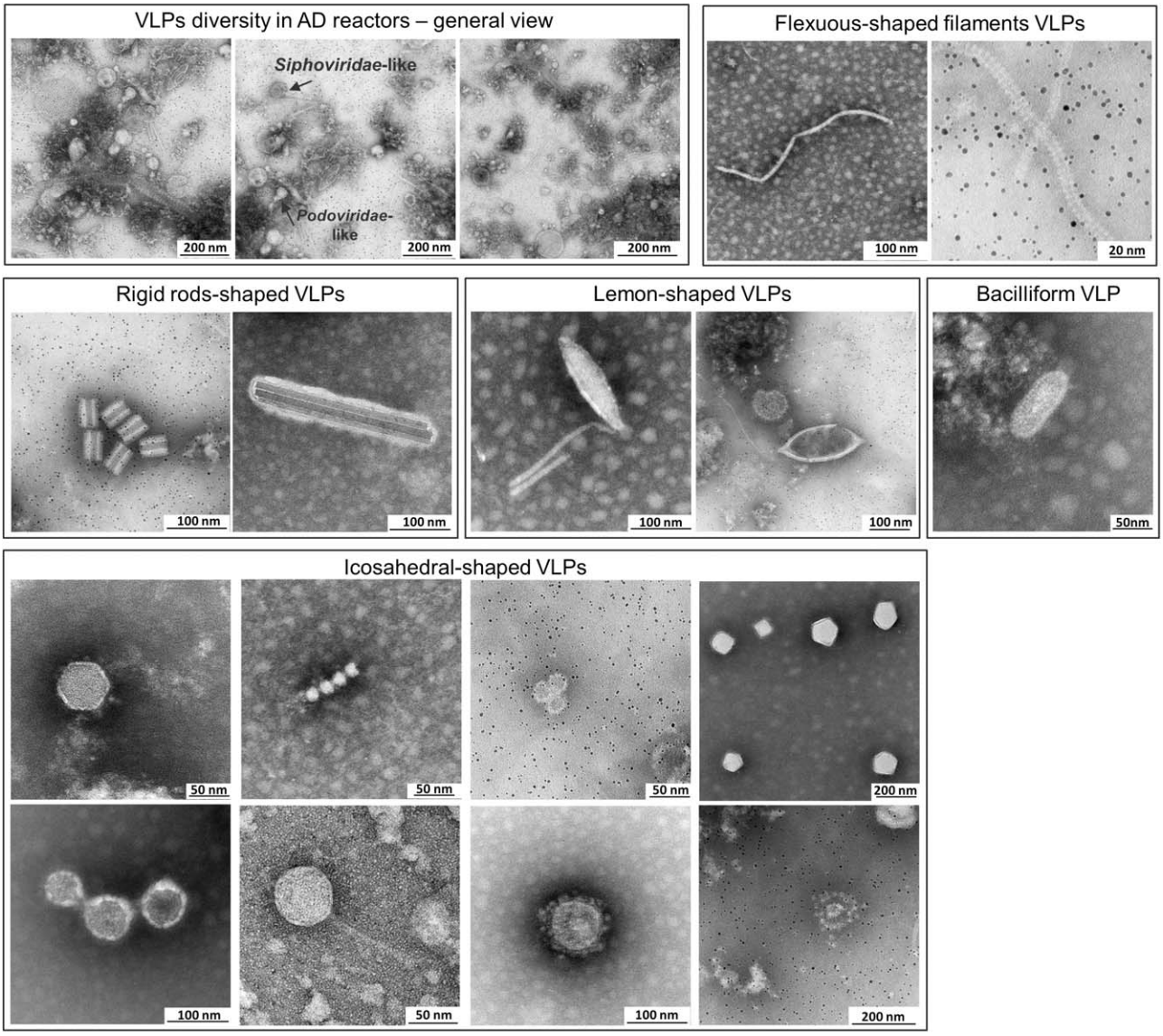

Fig. 1. Morphological diversity of VLPs isolated from sludge samples from AD reactors studied by TEM. Representative samples for each reactor type were studied, including waste water treatment plants (sample S3), farm digesters (sample B10) and biogas units treating a mixture of municipal solid waste and agro-food residues (sample F4).

range of the longest contigs identified in this study. However, further sequence-independent (e.g. based on the calculated contigs spectra, thus not biased by existing databases) contig binning (identification of contigs derived from one genome) should help to re-construct additional new viral genomes, as it was reported for a novel crAssphage discovered in human faecal metagenomes (Dutilh *et al*., 2014). Out of 63 contigs that formed closed circles, only six had significant nucleotide similarity (BLAST bitscore above 50) to any vial sequence in Refseq virus database.

Co-assembly of the 10 RNA viromes (sample MK18 was discarded from the analysis due to insufficient concentration of the generated library) resulted in 1142 contigs, out of which only 191 were taxonomically assigned of viral origin, and 144 were classified as RNA viruses. The two largest RNA contigs were above 8 kb and putatively represent nearly complete genomes of novel ssRNA(+) *Picornavirales* (Fig. 3, Table S4). Contig_592 encodes for a putative polyprotein and shows only 21% of sequence similarity (query coverage of 39%) with *Heterosigma akashiwo* RNA virus (*Marnaviridae/ Marnavirus*). The three predicted ORFs for the second largest contig_416 encode three different polyproteins. The first two ORFs encode the non-structural proteins involved in replication, while the last one is a putative capsid polyprotein. They share 81–90% of sequence identity with the *Drosophila C* virus

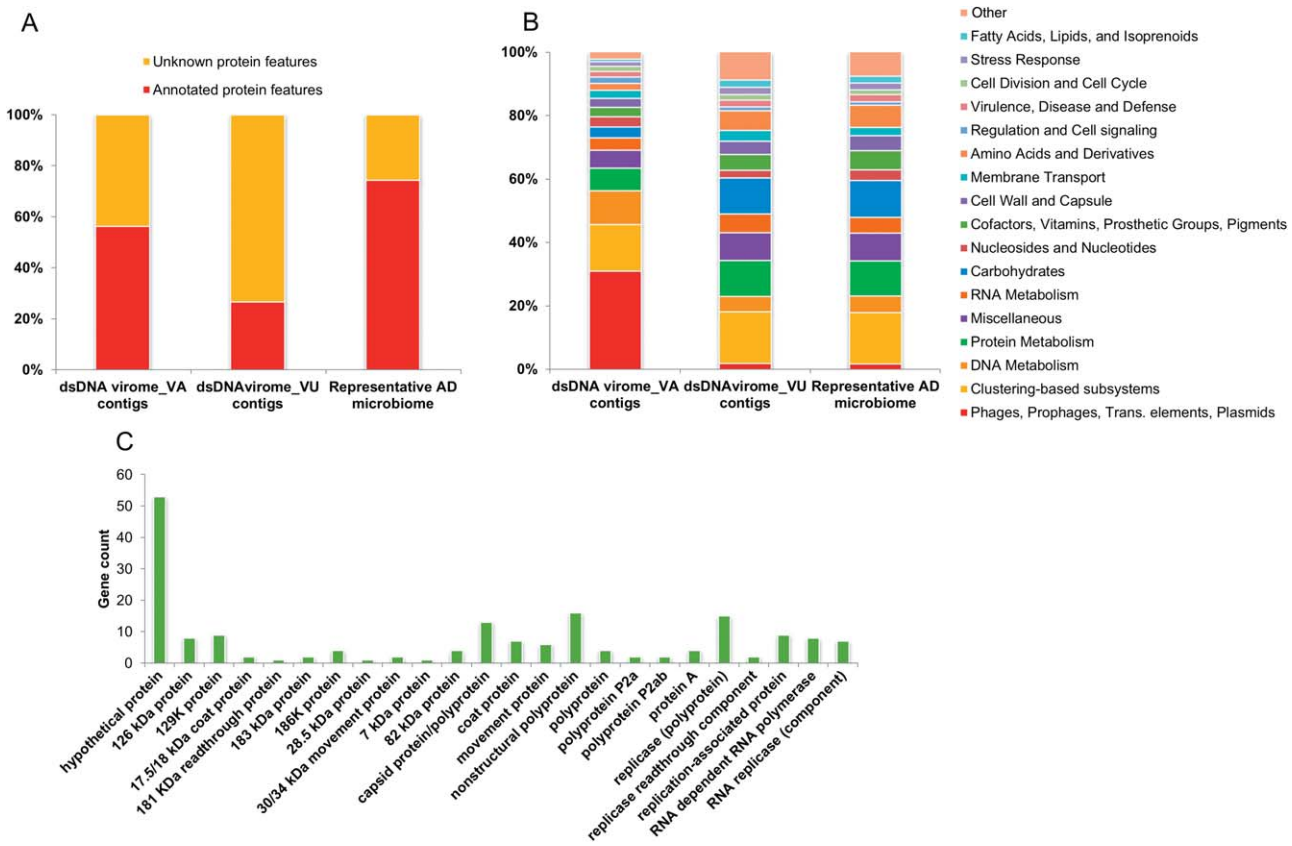

Fig. 2. Functional profiles for dsDNA and RNA viromes. (A) dsDNA virome all predicted protein features and (B) annotated protein features assigned to the different subsystems categories; dsDNA virome_virus-assigned (VA) contigs, dsDNA virome_virus-unassigned (VU) contigs and a representative AD microbiome. Annotated features for RNA virome virus-assigned contigs (C).

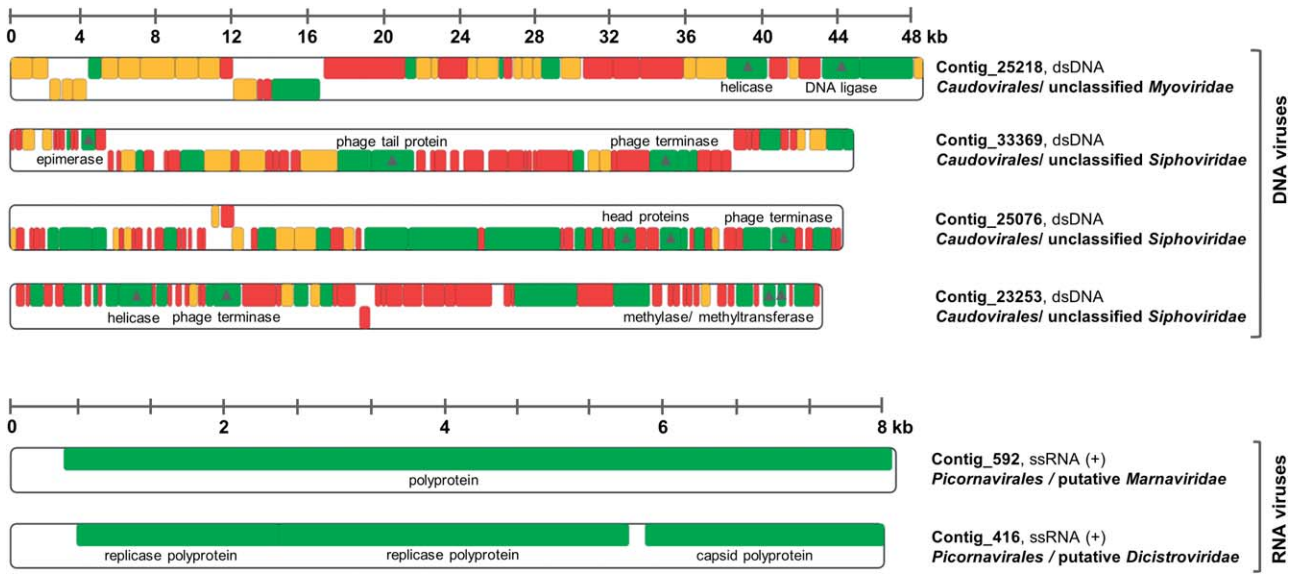

Fig. 3. The longest DNA and RNA contigs representing potentially complete viral genomes. Colour code for the predicted genes stands for genes affiliated to Refseq virus (green), PFAM only affiliation (yellow) and uncharacterized genes (red). Selected typical phage genes are indicated (▲).

(*Dicistroviridae/Cripavirus*), for the query coverage of 94–100%. In general, most of the predicted protein features for virus-assigned RNA virome contigs encode for either hypothetical proteins, large polyproteins or proteins annotated according to their size, e.g. putative 126 kDa protein. Therefore, they could not be assigned to functional categories using the subsystems approach (Fig. 2C).

dsDNA viromes are more diverse than RNA viromes in AD reactors

Following the report of Allen and colleagues (2013), showing that a commonly used tool PHACCS (Angly *et al*., 2005) tends to underestimate viral richness in environmental samples, in our study the estimation of viral richness was done using CATCHALL (Bunge *et al*., 2012) and normalized contigs spectra. Only contigs that were annotated of putatively viral origin were used to calculate the richness. On the one hand, this conservative approach most probably resulted in diversity indices that were minimum estimates, and any novel viruses that show little or no sequence similarity to the previously sequenced ones were not considered. On the other hand, restriction of the analyses to contigs containing ORFs with significant similarity to previously sequenced viral genes should help to reduce the overestimation of richness caused by free contaminant bacterial gDNA (Roux *et al*., 2013). Spurious singletons (e.g. sequences that could not assemble in larger contigs) that can easily inflate species richness were further addressed by applying a discounted model (Allen *et al*., 2013). As a result, the estimated richness of DNA viruses was in general higher in farm digesters and biogas units treating a mixture of solid waste and agro-food residues in comparison to WWTP (Table 2). Moreover, thanks to the normalization approach chosen in our study, we could quantitatively compare the richness between the dsDNA and RNA viruses. Accordingly, the estimated richness of dsDNA viruses was outnumbered by at least two orders of magnitude that of RNA viruses present in the analysed AD reactors.

dsDNA viromes in AD reactors are dominated by head-tailed viruses

The taxonomic analysis of the 27 K contigs assigned of viral origin indicated the recovery of predominantly dsDNA phages (94.6%). In total for dsDNA viromes 21 different families were assigned; however, the number varied between 11 and 17 according to the sample origin. Around 5% of identified viruses could not be classified to the existing viral families/orders. Head-tailed phages of the order *Caudovirales* constituted 77% of all the viruses identified in the AD reactors studied with three families, *Myoviridae*, *Siphoviridae* and *Podoviridae*, that accounted here, respectively, for 13.2%, 51.5% and 6.9% of all the dsDNA viruses (Fig. 4). Other dominant viral groups were the members of the family *Tectiviridae* (6% of all the DNA viruses identified), followed by *Mimiviridae* (1.6%) and *Phycodnaviridae* (1%), with the two former families being known to contain giant viruses infecting eukaryotes (Claverie *et al*., 2009).

On the one hand, the dominance of *Caudovirales*-related sequences is consistent with a previous study, where the *Caudovirales* order represented approximately 93.8% of identifiable viral sequences at different stages of typical waste water treatment systems in the US (Tamaki *et al*., 2012). On the other hand, *Caudovirales* genomes account for nearly 89% of all sequenced viral genomes (NCBI Viral Genomes, August 2015), biasing current databases and what may influence their predominance in metagenomic studies. Nevertheless, according to the recent revision of the electron microscopic descriptions of prokaryotic viruses, *Caudovirales* represent the vast majority (over 96.3%) of all known dsDNA bacteriophages (Ackermann and Prangishvili, 2012), supporting the reasonable assumption that *Caudovirales* might indeed dominate the virome of anaerobic reactors.

Prevalence of lytic ssDNA viruses (e.g. *Microviridae*), which play a significant ecological role by contributing to the bacterial cell lysis (Suttle, 2007), was documented for some soil and aquatic habitats (Krupovic *et al*., 2011), and was recently revised for ocean samples (Labonte and Suttle, 2013). Nonetheless, due to the sequencing approach undertaken in our study, ssDNA viruses were not targeted; therefore, it is not known if and at what quantities they exist in AD reactors. For this reason, to draw more conclusions on how phages interact with their microbial hosts in AD environments, it is important to further analyse ssDNA viruses to have a complete view of viral assemblages. Moreover, the monitoring of temporal changes of microbial communities and the corresponding viral assemblages would enable investigation of the rate of phage-mediated bacterial and archaeal cell turnover, thus to assess the 'phage kills the winner' strategy in a methanogenic world.

Contigs assigned to known archaeal viruses are present in AD reactors

DNA contigs similar to known Euryarchaeal (*Siphoviridae*, *Salterprovirus*) and Crenarchaeal (*Fuselloviridae*, *Bicaudaviridae*, *Lipothrixviridae*, *Rudiviridae*) viral families were detected in analysed sludge samples. Until now, only five methanogen-infecting viruses have been characterized, and they belong to either *Siphoviridae* or *Myoviridae* families (Pietila *et al*., 2014). Here, roughly 16 contigs were of putative *Siphoviridae* origin, and they

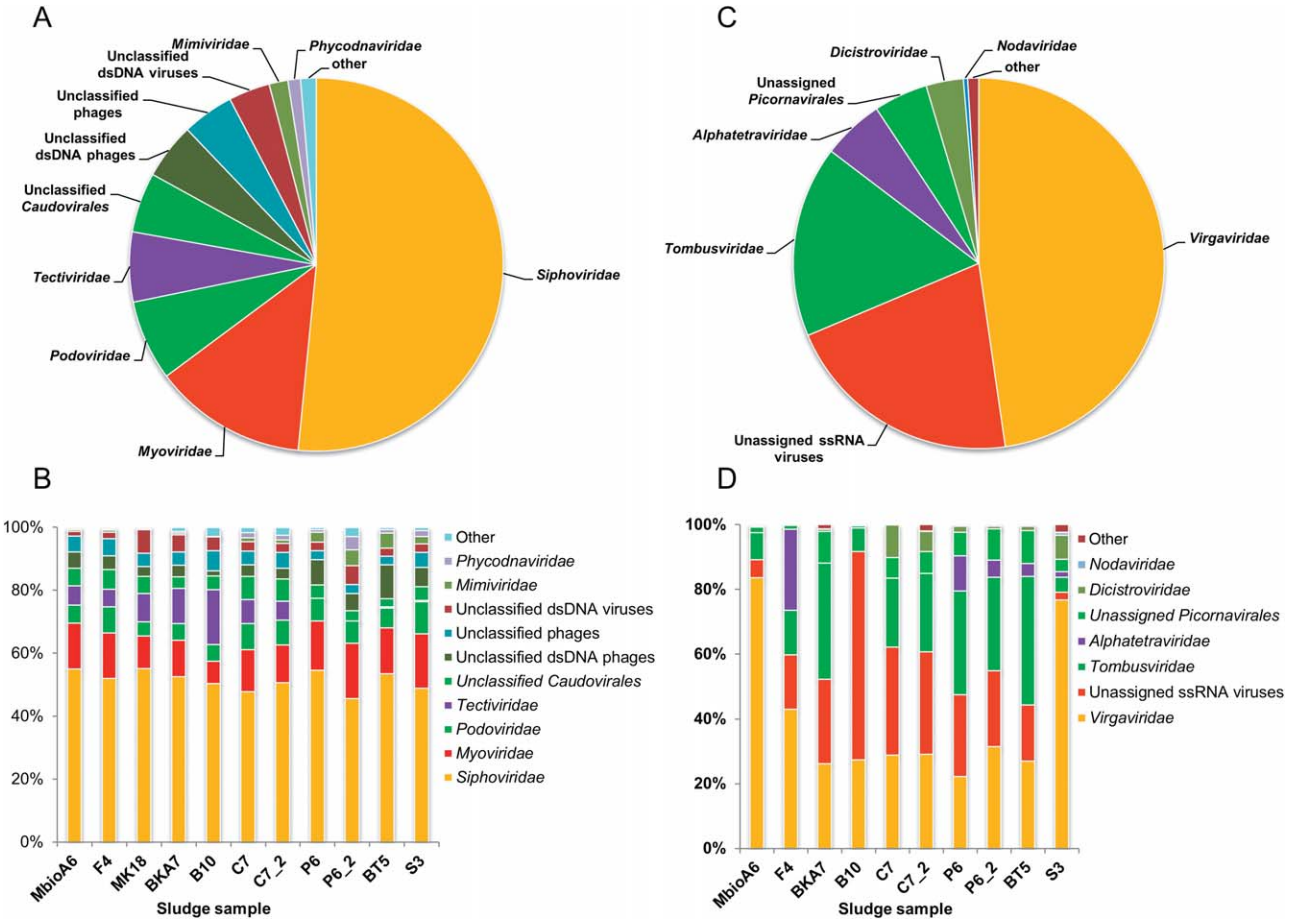

Fig. 4. Taxonomic characterization at the family level of dsDNA (A and B) and RNA (C and D) viromes. Upper pie charts show a global representation of dsDNA and RNA viral families for all sludge samples from AD reactors. Lower bar diagrams show a sample-specific representation of viral families observed.

were most closely related to Archaeal BJ1 virus-like. Interestingly, 19 contigs of different size range (643–4461 bp) were putatively assigned to *Salterprovirus*, and each of the contigs encoded type-B DNA polymerase, which is a signature of this viral genus. Remarkably, at least two different genotypes of *Salterprovirus*-like viruses were present, one prevailing in WWTP AD reactors and the other in farm and municipal solid waste treating reactors. Until now, two genomes of *Salterprovirus* were sequenced, namely His1 and His2, and the viruses were shown to infect only the extremely halophilic genus *Haloarcula* (Bath *et al*., 2006). Multiple other dsDNA contigs were assigned to unclassified archaeal viruses that were mainly similar to *Haloviruses*, *Pyrococcus abyssi* virus 1 and Hyperthermophilic Archaeal Virus 2. With regard to Crenarchaeal viruses, the highest number of contigs was affiliated with the *Bicaudaviridae* family.

The presence in AD reactors of numerous contigs associated with viruses previously discovered in hypersaline or hyperthermophilic environments suggests that despite the widespread presence of archaea on our planet, specific analysis of archaea-infecting viruses has only been done in a few, usually extreme environments (Prangishvili *et al*., 2006). Indeed, for over 6000 bacterial viruses studied to date, less than 100 archaeal viruses have been described, and most of them infect extremophiles (Ackermann and Prangishvili, 2012). Moreover, except for the head-tailed viruses affecting archaea, most of the characterized archaeal viruses share genome sequence similarity only with the taxonomically closely related viruses (Pietila *et al*., 2014). Therefore, the identification of novel archaeal viruses that relies on a comparative analysis with sequences deposited in public databases (66 complete genomes were deposited in NCBI Viral Genomes, August 2015) does not correctly reflect their diversity in newly studied natural environments; hence, the genetic diversity of archaeal viruses still remains poorly characterized.

RNA viromes are dominated by plant-infecting viruses

Unlike the pool of dsDNA viruses in AD reactors, which were composed mainly of bacteriophages, RNA viromes, assigned to a total of 14 different families, contained mostly viruses that infect eukaryotes. In general, RNA viruses infecting bacteria are rare, while up to now there were no archaeal RNA viruses reported (Nasir *et al*., 2014). In consequence, except for one contig assigned as unclassified *Levivirus Acinetobacter phage AP205*, the two known families of bacterial RNA viruses, namely dsRNA *Cystoviridae* and ssRNA *Leviviridae*, were not detected in AD reactors. Interestingly, 46% of virus-assigned RNA contigs were potential plant pathogens and were associated with five ssRNA(+) viral families, namely *Virgaviridae*, *Betaflexiviridae*, *Tombusviridae*, *Potyviridae* and *Secoviridae* (Table S5). Based on the identity scores to viral genomes listed in Refseq, 11 known viruses were identified using the criteria of ≥ 80% sequence identity over ≥ 90% of sequence length. Several contigs represented *Tobamovirus* genus (*tobacco mosaic* virus group), including many members that are widespread and well-characterized plant pathogens, e.g. cucumber green mottle mosaic virus, pepper mild mottle virus, turnip vein-clearing virus, etc. To our surprise, WWTP and municipal solid waste treating AD reactors were the main reservoirs of these viruses with, respectively, nine and ten *Tobamovirus* members present, while only six were found in farm AD reactors. The prevalence of plant pathogenic viruses in WWTP AD reactors (between 59 and 97% of RNA contigs annotated as plant pathogens were present in the sludge samples from WWTP AD reactors analysed here) could be explained by their dominance in human faeces which are input substrates. The contaminated food was proposed to be the source of plant-infecting RNA viruses in human faeces (Zhang *et al*., 2005).

Although the diverse plant pathogenic viruses were common in RNA viromes of different AD reactors, the values are not quantitative on an absolute basis; thus, reliable estimation of their absolute abundance cannot be performed. Nevertheless, assessment of their quantities and the viability would be a logical extension of this study, as the residue of the AD process, namely the digestate, is frequently used as a fertilizer in agriculture. Additionally, the full chain of reactors (fed reactors, post-digesters and storage tanks) should be studied to evaluate the potential of the AD cascade to reduce plant pathogenic viruses. Moreover, assessing plant pathogenic viruses in sludge from WWTPs through metagenomics could prove a useful tool to trace the human-mediated dissemination of plant viruses of potential quarantine significance to new areas of the world (Candresse *et al*., 2014).

Human and animal pathogenic viruses are seldom in methanogenic reactors

Around 0.68% of annotated DNA contigs were tentatively classified into three main viral families (*Adenoviridae*, *Herpesviridae and Poxviridae*) that include potential human pathogens (Table S4). However, except for three contigs that were putatively annotated as human herpes viruses, none of the other DNA contigs were directly assigned as human pathogens. Whereas our study did not reveal the presence of human adenoviruses in the AD reactors of the Schifflange WWTP [results confirmed by an additional quantitative polymerase chain reaction (qPCR) assay, Table S6], an integrated cell culture PCR approach showed their presence in the main influent and effluent streams of the aerobic treatment tanks of the same plant (Ogorzaly *et al*., 2013). Since adenoviruses represent risk to human health, their removal through AD treatment is an alternative that could be considered to re-design modern WWTPs. Nevertheless, as human adenoviruses were detected in influent and effluent sludge from other mesophilic AD reactors located in the US (Bibby and Peccia, 2013a,b), a global monitoring of different WWTPs should be performed with a standardized procedure (e.g. excluding the MDA step that biases the final quantitative results) to assess the efficiency of the removal of human viral pathogens through the AD treatment. No annotated human gastrointestinal RNA viruses, e.g. rotaviruses, astroviruses, certain members of coronaviruses, etc., were detected in the AD reactors studied. Also, previous metagenomic analysis of human faeces from healthy individuals indicated that less than 2.9% of sequencing RNA reads resembled animal viruses (Zhang *et al*., 2005), meaning that although the raw sewage represents the effluent from urban human activities, provided that a large part of the population is composed of healthy human individuals, human gastrointestinal viruses are uncommon in methanogenic reactors of WWTPs.

Putative animal-infecting viruses represented 1.12% of annotated contigs, with the majority being assigned to five viral families/ orders, including *Poxviridae* (mostly *Entomopoxvirinae*), *Ascoviridae*, *Iridoviridae, Herpesvirales* and *Baculoviridae*. Except for *Herpesvirales*-assigned contigs, most of the other contigs were tentatively assigned to insect-infecting viruses, with *Lepidoptera* moths (many species are common pests on crops) being identified as main hosts.

Amoeba-infecting giant-like viruses are widespread in methanogenic reactors

Acanthamoeba polyphaga mimivirus was the first discovered giant virus, isolated from an amoebal co-culture, and classified as a new family *Mimiviridae*-'mimicking the

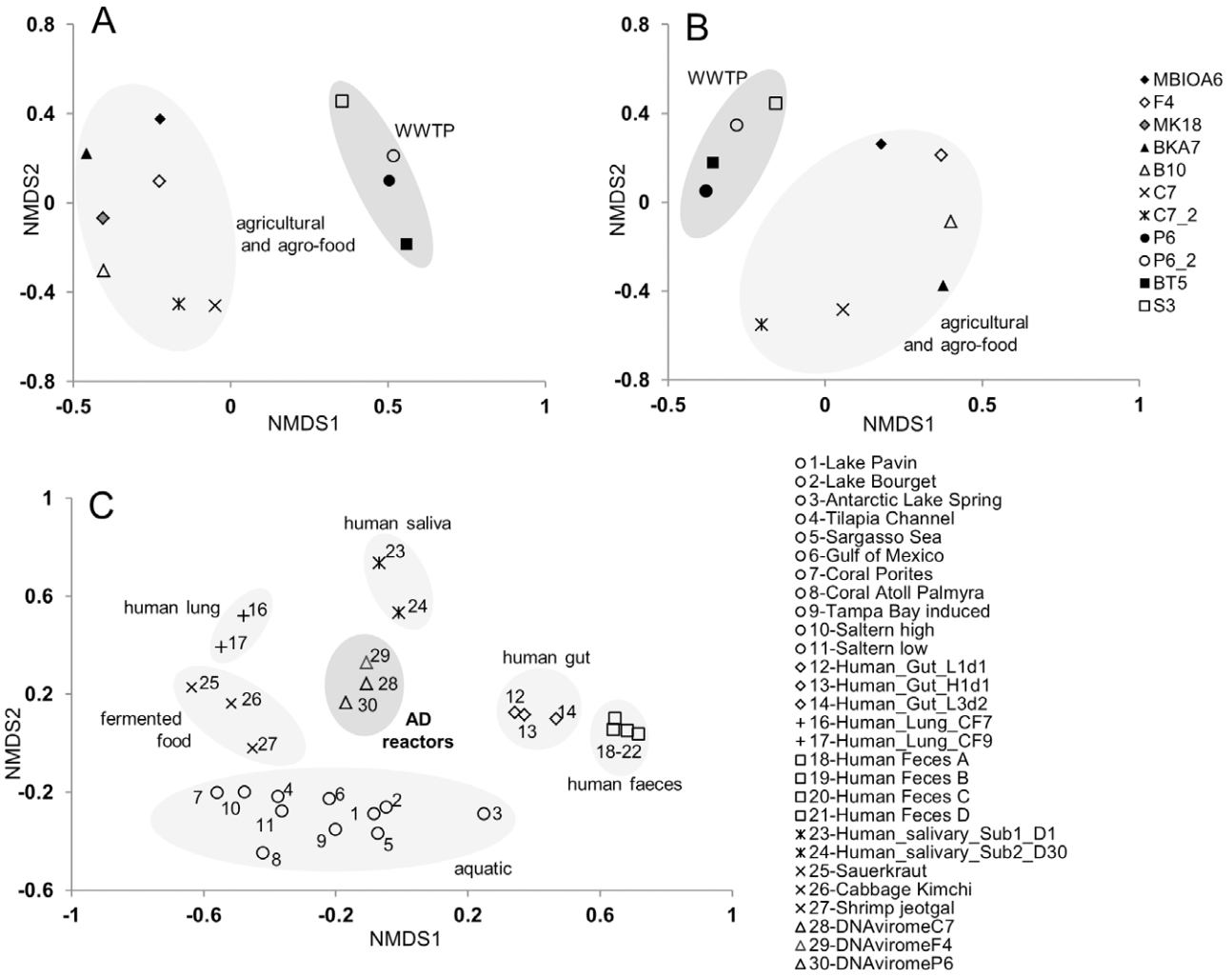

Fig. 5. NMDS analysis of dsDNA (A) and RNA (B) viromes. The comparison of representative dsDNA viromes analysed in this study with 27 other selected environmental DNA viromes, including different aquatic habitats, human faeces, gut, saliva and lungs and fermented food (C). The lowest stress value for A and B was 0.2. For C, the MDS coordinates were calculated with METAVIR2, and the stress value was given as 7.69% with the R^2 non-metric fit of 0.994.

microbe' (Scola *et al*., 2003). Amoebas, the main hosts of these viruses, have been isolated from various environments, including sewage treatment systems; therefore, not surprisingly approximately 6% of annotated contigs, mainly from waste water AD reactors, were tentatively assigned to different giant viruses, e.g. *Acanthamoeba polyphaga mimivirus*, *Megavirus chilensis*, *Acanthamoeba polyphaga moumouvirus* and *Pandoravirus dulcis*. Giant viruses are dsDNA viruses with an extraordinary diameter of around 700 nm and a genome size of 1.2 Mb. Recent studies have brought even more attention to these viruses as they indicated that in addition to amoebas, also vertebrates, including humans, may be hosts of these viruses (Abrahao *et al*., 2014).

Comparative analysis of viromes demonstrates the difference in viral diversity between the WWTPs and other methanogenic reactors versus other environments

As one of the goals of this study was to determine if the viral assemblages from different methanogenic reactors were similar, the different dsDNA and RNA viromes were compared by applying non-metric multidimensional scaling (NMDS) on annotated contigs spectra. Consistent with the NMDS results, dsDNA and RNA viromes from WWTPs were clearly distinct from the farm digesters and organic-waste treating reactors ($P < 0.001$), while farm and organic-waste AD reactors clustered together (Fig. 5A and B).

Moreover, further BLAST-based comparative genomic analysis (to avoid the bias from the contaminant bacterial gDNA, only contigs assigned to be of viral origin were considered) and conducted in METAVIR2 (Roux *et al*., 2014b) revealed that the taxonomic composition of DNA viral assemblages in AD reactors was different from 27 other selected environments, including human gut, saliva, lungs and faeces, fermented food, deep sea sediments, and other aquatic environments (Fig. 5C). These observations suggest that viruses in AD reactors, mainly dsDNA head-tailed phages, form assemblages that are different from those found in other environments. Indeed, at the genomic level, *Caudovirales* are an extremely divergent group with mosaic genomes composed of conserved and variable regions. Most of their homologous nucleotide and protein sequences often display none or only 10–20% identity in pairwise comparisons (Krupovic *et al*., 2011), essentially leading to a high number of hypothetical proteins discovered in viral metagenomes (Hurwitz and Sullivan, 2013). Moreover, it has been assessed at a global scale that between 10^9 and 10^{10} fully functional bacteriophage genomes are produced via non-homologous recombination every second (Hendrix, 2009), and once they are viable they can easily spread within the population of closely related viruses, thus keeping on generating diversity. As phages obligatorily require their bacterial hosts, and the concentration and diversity of bacteria are high in AD reactors, no wonder that extensive homologous or non-homologous re-combinations, genomic segment duplications, etc. may take place, leading to a rapid evolution and an immense and novel diversity of bacteriophages in these environments.

Conclusions

In spite of the fact that phages are recognized to play key roles in microbial ecology by controlling strain composition and abundance, e.g. some of the known lysogenic archaeal viruses like *Pleolipoviridae* were shown to cause persistent infections resulting in a continuous virus production and host growth retardation (Pietila *et al*., 2014), until now viruses remain largely unstudied in methanogenic digesters. Moreover, the main interest to study viruses in this environment is not only the fact that by shaping microbial dynamics they can negatively influence the process of methane production, but also the development of a phage-mediated treatment of anaerobic sludge, by regulating the abundance of key functional microbial groups, has the potential to control some of the process problems, like foaming, sludge dewaterability and digestibility, or the competition between nuisance and functionally important bacteria (Withey *et al*., 2005). Although current sequencing methodologies, e.g. metagenomics, enable the sequencing of wild viral assemblages with unprecedented depth and resolution, due to an incredibly high diversity of viral genomes, the majority of viral sequences are not yet found in databases, which prevents their complete characterization when relying on sequence homology-based identification. Therefore, further genetic analysis of gene contents of the identified phage contigs should provide more information about their lifestyle, and investigations with methods such as viral tagging (Deng *et al*., 2012) or digital PCR (Tadmor *et al*., 2011) will be required to expand our knowledge on the interactions between viruses and their hosts in AD environments.

Experimental procedures

AD reactors and sampling procedure

After mixing the reactor to collect a representative sample of 10–100 l, depending on the biogas unit, a sludge volume of 50 ml was snap frozen on site and kept at –80°C until further analysis. All reactors sampled in this study were the fed reactors. Metabolic parameters, including pH, total inorganic carbon, total ammonium-nitrogen, and total solid and volatile solid contents, were determined for each AD reactor at the time of sampling. Total solids (TS, 24 h at 105°C) and volatile solids (VS, 6 h at 550°C) were determined according to the 4630 Voreinigen Deutsche Ingeniören (VDI) norm. Total inorganic carbon and ammonium-nitrogen concentration (NH_4-N) were measured in conformity with the manufacturers' protocol, using the BiogasPro system (RIMU, Königsbrunn, Germany).

Filtration-based extraction of VLPs

Sludge samples were diluted with an SM buffer to 0.5 l, implemented with 5–10 stainless milling beads (Φ5 mm) and shaken vigorously (250 r.p.m. min^{-1}) for 30 min to disperse flocks and to detach the biomass and viruses from vegetal fibres, then centrifuged at 4000 *g* for 20 min to sediment large-size particles and bacteria. The resulting supernatant was subjected to a three-step pre-filtration with an 8 μm Sartopure nitrocellulose filter (Sartorius) placed in a Büchner funnel under vacuum, a 5 μm pore Minisart syringe filter (Sartorius) and finally with a 0.8 μm MF-Millipore membrane filter placed in a Büchner funnel. The filtrate was applied onto a TFF system (Sartorius) mounted with a Sartocon Slice 200 Hydrosart cassette of 0.2 or 0.45 μm. The obtained permeate was ultracentrifuged at 45 000 *g* for 3 h at 4°C, and the resulting pellet was re-suspended in 2 ml of SM buffer.

Determination of the efficiency of the VLPs extraction procedure

A representative sludge sample was autoclaved and spiked in with 1 ml of ΦX174 phage suspension corresponding to a concentration of 5×10^8 pfu ml^{-1}. The sample was thoroughly shaken and subjected to all of the pre-filtration and TFF steps as described above. At each step, an aliquot was retained and a plaque-forming assay was applied in accordance with

ISO 10705-2 to evaluate the phage recovery. Briefly, the fractions collected in check-points 0-8 (Table S1) were serially diluted (dilutions from 10^{-1} to 10^{-7}), and 1 ml of each dilution was mixed with 1 ml of warm semi-solid Modified Scholtens Agar (ssMSA) medium and 1 ml of *Escherichia coli C* culture (ATCC 13706, OD = 0.287) in Modified Scholtens Broth (MSB) medium. Afterwards, the mixture was poured over a plate containing Modified Scholtens Agar (MSA) agar to create a second layer. For the positive control, the same procedure was performed with 1 ml of ΦX174 phage stock suspension (10^2 pfu ml^{-1}), and for the negative control only bacteria in MSB and ssMSA media were used to create a second layer. After an overnight incubation at 37°C, plaques formed on the agar were counted and the concentration of phages was calculated according to the following formula:

$$X = \frac{N}{(n_1V_1F_1)+(n_2V_2F_2)+(n_3V_3F_3)}$$

where X = concentration of phages in pfu ml^{-1}, N = total number of plaques, n_1 = number of replicates per dilution, V_1 = volume used, and F_1 = dilution factor.

TEM

For a selected number of samples, but including each sludge origin, VLPs from 25-fold concentrated stocks were deposited on formvar-coated copper grids, stabilized with evaporated carbon film (200 mesh, S162 Plano GmbH) and negatively stained with either 2% (wt vol^{-1}) uranyl acetate or 2% phosphotungstic acid. The air-dried grids were examined with a LEO 922 OMEGA (Carl Zeiss) TEM microscope operated at 120 kV.

Preparation of viral nucleic acids

Following the extraction of VLPs, 200 µl of viral concentrate was treated with 8 U of Turbo DNase (2 U μl^{-1}, Ambion) for 30 min at 37°C to remove extracellular genomic DNA. Viral DNA and RNA were extracted simultaneously with a QIAampMinElute Virus Spin Kit (QIAGEN), according to the manufacturer's instructions, except for a carrier RNA which was omitted from the protocol. Following the extraction, each sample was divided into two parts and treated with either RNase A (Sigma) at a final concentration of 0.2 mg ml^{-1} or Turbo DNA-free kit (Ambion), and following the manufacturer's instructions, to obtain respectively viral DNA and RNA. Concentrations of viral DNA and RNA were determined accordingly with Qubit dsDNA HS assay and a Qubit RNA HS assay and were in the range of 19.03–0.21 ng μl^{-1} for DNA and 0.3–0.5 ng μl^{-1} for RNA respectively. No amplification of total viral nucleic acids was performed, as the quantity of purified viral DNA and RNA was sufficient for preparation of metagenomic libraries.

Preparation of viral dsDNA and RNA metagenomic libraries and sequencing

dsDNA libraries were prepared with a Nextera XT DNA sample preparation kit (Illumina), according to the manufacturer's instructions, except for the normalization step. Instead, the quantity of each library was determined by a qPCR with a Kapa Sybr Fast universal qPCR kit (Kapa Biosystems), and the libraries were mixed in equivalent molar concentrations prior to sequencing. A SMARTer Stranded RNA-seq kit (Clontech Laboratories) was used according to the manufacturer's instructions to prepare viral RNA libraries. Again, each library was quantified with a qPCR and pooled in equimolar concentrations. DNA and RNA libraries were sequenced separately on MiSeq (Illumina) in three sequencing runs, using MiSeq Reagent Kits V2-300, V2-500 and V3-600.

Bioinformatic analysis of DNA and RNA viromes

The software CLC Genomics Workbench 7.0 was used for reads filtering, assembly and mapping. Following trimming for adapters, quality scores and read length, 42 M (dsDNA virome) and 13 M (RNA virome) reads were used to assemble into contigs. The summary of these steps is given in Tables S2 and S3. The two *de novo* co-assemblies were prepared, which incorporated sequence reads that passed the quality control from all samples (dsDNA and RNA viromes separately). To determine the relative abundance and occurrence of virus-like sequences in annotated contigs, contigs spectra were prepared. Briefly, reads were de-replicated and mapped back to the *de novo* assembled contigs for dsDNA and RNA viromes separately, and the average abundance was calculated as DNA-RPKM. The obtained annotated contigs abundance spectra were used to calculate alpha diversity with CatchAll version 3.0, as implemented into Mothur v.1.31.1 (Schloss *et al.*, 2009) and rarefaction curves. CatchAll, by comparing a suite of flexible parametric and non-parametric models, returns the best model according to statistical and heuristic criteria. Moreover, by statistically discounting low-frequency observations, it provides the discounted richness estimation as well (Bunge *et al.*, 2012; Allen *et al.*, 2013). Similarity in community membership between the different DNA and RNA viromes was compared by applying NMDS on annotated contigs spectra and using traditional Jaccard similarity coefficient based on the observed richness.

Assembled contigs were taxonomically assigned using a BLAST comparison with the Refseq complete viral genomes protein sequences database from NCBI (release from 2014-03-13) using BLASTP (threshold of 50 on the BLAST bitscore) as implemented in Metavir2 (Roux *et al.*, 2014b). Contigs with multiple ORFs that showed highest similarity to viral genes encoded by distinct genomes were taxonomically assigned according to the best BLAST bitscore. The raw sequence data have been submitted to the NCBI Sequence Read Archive under Accession Number SRP051200. Assembled and annotated contigs are available via Metavir2 project Accession Numbers 3506 (dsDNA virome) and 3518 (RNA virome). Contigs classified of viral origin and those that could not be classified as viruses were further functionally annotated using the MG-RAST server (Meyer *et al.*, 2008), and are accessible thought the MG-RAST ID Numbers 4559444.3 and 4571041.3 respectively.

Real-time quantitative PCR (qPCR)

Hexon-specific adenoviral primers AdF and AdR and a hydrolysis probe AdP1 (Hernroth *et al.*, 2002) were used to

amplify the major capsid protein coding gene to detect human adenovirus serotypes A, C, D, E and F. The PCR amplifications were performed in 20 μl volume reactions including 1x TaqMan Environmental Master Mix 2.0 (Applied Biosystems), each primer at the final concentration of 0.9 μM and probe at 0.225 μM. The PCR cycling conditions were as follows: 95°C for 10 min, followed by 50 cycles of 95°C for 15 s and 60°C for 1 min. A standard curve was used for absolute quantification.

Acknowledgements

This work was supported by the Fonds National de la Recherche, Luxembourg (FNR CORE 2011 Project GASPOP, C11/SR/1280949: Influence of the Reactor Design and the Operational Parameters on the Dynamics of the Microbial Consortia Involved in the Biomethanation Process). We thank Cécile Walczak and Leslie Ogorzaly from the ERIN department of our institute for their help and suggestions during the VLPs filtration and TEM preparations. We thank Bénédicte De Vos, Anaïs Noo and Marie Fosséprè for their technical support. The authors are most thankful to the members of the Biogas-Vereenegung ASBL, to François Delion (SOVACOM sarl), the members of the ASBL Au pays de l'Attert, and the partnership of the Ecobiogaz project of the INTERREG IV A programme for providing a year-round access to the biogas plants and for their help in sampling the reactors for sludge. Waste water plant managers and their respective technical staff are also warmly acknowledged for their contribution in sampling the reactors on their plants: André Detaille, Jean-Jacques Schmit and Régine Reinert (STEP, Bettembourg), Gerry Bissen and Gianni Di Pentima (SIVEC, Schifflange), Piero Daresta, Roland Troes, and Raymond Erpelding (SIACH, Pétange). We kindly thank Jillian M. Lenné for English correction. We would also like to thank the editor Dr Victoria Orphan and the anonymous reviewers for their critical revision that helped to improve this manuscript. The authors have declared that no competing interests exist.

References

Abrahao, J., Dornas, F., Silva, L., Almeida, G., Boratto, P., Colson, P., *et al*. (2014) *Acanthamoeba polyphaga* mimivirus and other giant viruses: an open field to outstanding discoveries. *Virol J* **11:** 120.

Ackermann, H.W., and Prangishvili, D. (2012) Prokaryote viruses studied by electron microscopy. *Arch Virol* **157:** 1843–1849.

Allen, H., Bunge, J., Foster, J., Bayles, D., and Stanton, T. (2013) Estimation of viral richness from shotgun metagenomes using a frequency count approach. *Microbiome* **1:** 5.

Angly, F., Rodriguez-Brito, B., Bangor, D., McNairnie, P., Breitbart, M., Salamon, P., *et al*. (2005) PHACCS, an online tool for estimating the structure and diversity of uncultured viral communities using metagenomic information. *BMC Bioinformatics* **6:** 1–9.

Bath, C., Cukalac, T., Porter, K., and Dyall-Smith, M.L. (2006) His1 and His2 are distantly related, spindle-shaped haloviruses belonging to the novel virus group, *Salterprovirus*. *Virology* **350:** 228–239.

Bibby, K., and Peccia, J. (2013a) Identification of viral pathogen diversity in sewage sludge by metagenome analysis. *Environ Sci Technol* **47:** 1945–1951.

Bibby, K., and Peccia, J. (2013b) Prevalence of respiratory adenovirus species B and C in sewage sludge. *Environ Sci Processes Impacts* **15:** 336–338.

Bunge, J., Woodard, L., Böhning, D., Foster, J.A., Connolly, S., and Allen, H.K. (2012) Estimating population diversity with CatchAll. *Bioinformatics* **28:** 1045–1047.

Candresse, T., Filloux, D., Muhire, B., Julian, C., Galzi, S., Fort, G., *et al*. (2014) Appearances can be deceptive: revealing a hidden viral infection with deep sequencing in a plant quarantine context. *PLoS ONE* **9:** e102945.

Chen, Y., Cheng, J.J., and Creamer, K.S. (2008) Inhibition of anaerobic digestion process: a review. *Bioresour Technol* **99:** 4044–4064.

Chien, I.C., Meschke, J.S., Gough, H.L., and Ferguson, J.F. (2013) Characterization of persistent virus-like particles in two acetate-fed methanogenic reactors. *PLoS ONE* **8:** e81040.

Claverie, J.M., Grzela, R., Lartigue, A., Bernadac, A., Nitsche, S., Vacelet, J., *et al*. (2009) Mimivirus and *Mimiviridae*: giant viruses with an increasing number of potential hosts, including corals and sponges. *J Invertebr Pathol* **101:** 172–180.

Deng, L., Gregory, A., Yilmaz, S., Poulos, B.T., Hugenholtz, P., and Sullivan, M.B. (2012) Contrasting life strategies of viruses that infect photo- and heterotrophic bacteria, as revealed by viral tagging. *MBio* **3:** e00373–12.

Dutilh, B.E., Schmieder, R., Nulton, J., Felts, B., Salamon, P., Edwards, R.A., and Mokili, J.L. (2012) Reference-independent comparative metagenomics using cross-assembly: crAss. *Bioinformatics* **28:** 3225–3231.

Dutilh, B.E., Cassman, N., McNair, K., Sanchez, S.E., Silva, G.G.Z., Boling, L., *et al*. (2014) A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nat Commun* **5:** 4498.

Hendrix, R.W. (2002) Bacteriophages: evolution of the majority. *Theor Popul Biol* **61:** 471–480.

Hendrix, R.W. (2009) Jumbo bacteriophages. In *Lesser Known Large dsDNA Viruses*. Van Etten, J. (ed.). Berlin Heidelberg: Springer, pp. 229–240.

Hernroth, B.E., Conden-Hansson, A.C., Rehnstam-Holm, A.S., Girones, R., and Allard, A.K. (2002) Environmental factors influencing human viral pathogens and their potential indicator organisms in the blue mussel, *Mytilus edulis*: the first Scandinavian report. *Appl Environ Microbiol* **68:** 4523–4533.

Hurwitz, B.L., and Sullivan, M.B. (2013) The Pacific Ocean virome (POV): a marine viral metagenomic dataset and associated protein clusters for quantitative viral ecology. *PLoS ONE* **8:** e57355.

Ignacio-Espinoza, J.C., Solonenko, S.A., and Sullivan, M.B. (2013) The global virome: not as big as we thought? *Curr Opin Virol* **3:** 566–571.

Krupovic, M., Prangishvili, D., Hendrix, R.W., and Bamford, D.H. (2011) Genomics of bacterial and archaeal viruses: dynamics within the prokaryotic virosphere. *Microbiol Mol Biol Rev* **75:** 610–635.

Labonte, J.M., and Suttle, C.A. (2013) Previously unknown and highly divergent ssDNA viruses populate the oceans. *ISME J* **7:** 2169–2177.

Marine, R., McCarren, C., Vorrasane, V., Nasko, D., Crowgey, E., Polson, S.W., and Wommack, E.K. (2014) Caught in the middle with multiple displacement amplification: the myth of pooling for avoiding multiple displacement bias in a metagenome. *Microbiome* **2:** 3.

Maurice, C.F., Bouvier, C., de Wit, R., and Bouvier, T. (2013) Linking the lytic and lysogenic bacteriophage cycles to environmental conditions, host physiology and their variability in coastal lagoons. *Environ Microbiol* **15:** 2463–2475.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M., *et al*. (2008) The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9:** 386.

Minot, S., Bryson, A., Chehoud, C., Wu, G.D., Lewis, J.D., and Bushman, F.D. (2013) Rapid evolution of the human gut virome. *Proc Natl Acad Sci* **110:** 12450–12455.

Nasir, A., Forterre, P., Kim, K.M., and Caetano-Anolles, G. (2014) The distribution and impact of viral lineages in domains of life. *Front Microbiol* **5:** 194.

Ogorzaly, L., Cauchie, H.M., Penny, C., Perrin, A., Gantzer, C., and Bertrand, I. (2013) Two-day detection of infectious enteric and non-enteric adenoviruses by improved ICC-qPCR. *Appl Microbiol Biotechnol* **97:** 4159–4166.

Overbeek, R., Begley, T., Butler, R.M., Choudhuri, J.V., Chuang, H.Y., Cohoon, M., *et al*. (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res* **33:** 5691–5702.

Park, M.O., Ikenaga, H., and Watanabe, K. (2007) Phage diversity in a methanogenic digester. *Microb Ecol* **53:** 98–103.

Pietila, M.K., Demina, T.A., Atanasova, N.S., Oksanen, H.M., and Bamford, D.H. (2014) Archaeal viruses and bacteriophages: comparisons and contrasts. *Trends Microbiol* **22:** 334–344.

Prangishvili, D., Forterre, P., and Garrett, R.A. (2006) Viruses of the archaea: a unifying view. *Nat Rev Microbiol* **4:** 837–848.

Roux, S., Krupovic, M., Debroas, D., Forterre, P., and Enault, F. (2013) Assessment of viral community functional potential from viral metagenomes may be hampered by contamination with cellular sequences. *Open Biol* **3:** 130160.

Roux, S., Hawley, A.K., Torres Beltran, M., Scofield, M., Schwientek, P., Stepanauskas, R., *et al*. (2014a) Ecology and evolution of viruses infecting uncultivated SUP05 bacteria as revealed by single-cell- and meta-genomics. *Elife* **3:** e03125.

Roux, S., Tournayre, J., Mahul, A., Debroas, D., and Enault, F. (2014b) Metavir 2: new tools for viral metagenome comparison and assembled virome analysis. *BMC Bioinformatics* **15:** 76.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., *et al*. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75:** 7537–7541.

Scola, B.L., Audic, S.P., Robert, C., Jungang, L., de Lamballerie, X., Drancourt, M., *et al*. (2003) A giant virus in *Amoebae*. *Science* **299:** 2033.

Shapiro, O.H., Kushmaro, A., and Brenner, A. (2010) Bacteriophage predation regulates microbial abundance and diversity in a full-scale bioreactor treating industrial wastewater. *ISME J* **4:** 327–336.

Suttle, C.A. (2007) Marine viruses-major players in the global ecosystem. *Nat Rev Microbiol* **5:** 801–812.

Tadmor, A.D., Ottesen, E.A., Leadbetter, J.R., and Phillips, R. (2011) Probing individual environmental bacteria for viruses by using microfluidic digital PCR. *Science* **333:** 58–62.

Tamaki, H., Zhang, R., Angly, F.E., Nakamura, S., Hong, P.Y., Yasunaga, T., *et al*. (2012) Metagenomic analysis of DNA viruses in a wastewater treatment plant in tropical climate. *Environ Microbiol* **14:** 441–452.

Thingstad, T.F., and Lingen, R. (1997) Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. *Aquat Microb Ecol* **13:** 19–27.

Withey, S., Cartmell, E., Avery, L.M., and Stephenson, T. (2005) Bacteriophages-potential for application in wastewater treatment processes. *Sci Total Environ* **339:** 1–18.

Wommack, K.E., and Colwell, R.R. (2000) Virioplankton: viruses in aquatic ecosystems. *Microbiol Mol Biol Rev* **64:** 69–114.

Wu, Q., and Liu, W.T. (2009) Determination of virus abundance, diversity and distribution in a municipal wastewater treatment plant. *Water Res* **43:** 1101–1109.

Zhang, T., Breitbart, M., Lee, W.H., Run, J.Q., Wei, C.L., Soh, S.W.L., *et al*. (2005) RNA viral community in human feces: prevalence of plant pathogenic viruses. *PLoS Biol* **4:** e3.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. Correlation of the data obtained for dsDNA viromes for the two pore size of 0.22 μm (samples P6, C7) and 0.45 μm (samples P6_2, C7_2) during TFF.

Fig. S2. dsDNA and RNA viral communities richness.

Table S1. Evaluation of the VLPs isolation procedure – a plaque forming assay.

Table S2. Summary of the sequencing results for dsDNA and RNA viromes.

Table S3. Summary of reads assemblies for dsDNA and RNA viromes.

Table S4. Characteristics of viral dsDNA and RNA contigs. Only contigs annotated of viral origin were included in the table. Full dataset is accessible via METAVIR2 accession numbers 3506 (dsDNA virome) and 3518 (RNA virome).

Table S5. Putative plant pathogenic RNA viruses identified in sludge samples from AD reactors.

Table S6. qPCR-mediated detection of human adenoviruses in sludge samples from AD reactors.