# Molecular characterisation of hepatitis A virus strains from water sources in South Africa

R. Saïd, M. Wolfaardt and M. B. Taylor

## ABSTRACT

Hepatitis A virus (HAV) strains found in selected South African (SA) surface waters were characterised to establish what HAV types are circulating in the environment, thus reflecting circulation in the surrounding communities. Surface water samples used for irrigation or domestic purposes, and water samples from outflow of wastewater plants were collected from six provinces. Viruses were recovered from the samples using a glasswool adsorption-elution method and then further concentrated using polyethylene glycol/sodium chloride precipitation. After automated nucleic acid extraction, samples were analysed for HAV by real-time reverse-transcriptase polymerase chain reaction. HAV strains were genotyped by nucleotide sequence analysis of the capsid gene VP1 and the VP1/P2B junction. HAVs were detected in 76% (16/21) of the surface water samples and in 37% (19/51) of the samples from the wastewater plants. Strains were characterised from 32 of the 35 samples and classified within genotype IB. The presence of genotype IB in the water sources confirms human faecal contamination. Hence these faecally-contaminated water sources may be a potential transmission route of HAV infection and potential source of contamination of irrigated fresh produce in SA.

**Key words** | genotypes, hepatitis A virus, irrigation water, South Africa

**R. Saïd** (corresponding author)
**M. Wolfaardt**
Department of Medical Virology,
Faculty of Health Sciences,
University of Pretoria,
Private Bag X323,
Arcadia 0007 Pretoria,
South Africa
E-mail: *sadrachida19@yahoo.fr*

**M. B. Taylor**
National Health Laboratory Service,
Tshwane Academic Division,
Pretoria,
South Africa

## INTRODUCTION

Hepatitis A virus (HAV) is the leading cause of acute viral hepatitis worldwide (Costafreda *et al.* 2006; Pintó *et al.* 2010). Infection with HAV can proceed with or without symptoms depending on the age of the individual at the time of the infection and the presence or not of an underlying chronic liver disease (Collier & Oxford 2006; Nainan *et al.* 2006; Pintó *et al.* 2010). Children younger than five years of age rarely present with symptoms, while in older children and adults the infection can be severe (Dotzauer 2008; Spradling *et al.* 2009). HAV is an enteric virus transmitted via the faecal-oral route either directly from person-to-person or indirectly through the ingestion of faecally contaminated food and water (Nainan *et al.* 2006; Pintó *et al.* 2010, 2012; Rodríguez-Lázaro *et al.* 2012). The virus may less commonly be transmitted through blood transfusion and blood products, the sharing of needles by drug abusers and sexual contact – particularly between male homosexuals (Nainan *et al.* 2006; Pérez-Sautu *et al.* 2011a). HAV has a long incubation period ranging from 15 to 50 days, during which time the virus is actively replicating in the liver and found in the blood and faeces (Dotzauer 2008; Pintó *et al.* 2012). An infected individual may shed up to $10^{11}$ HAV particles per gram of faeces just before the onset of symptoms and it is at that point that the individual is most infectious (Pintó *et al.* 2010). A high and long-lasting viraemia was also identified for HAV occurring from before the beginning of symptoms to an average of 6 weeks after symptoms have started (Bower *et al.* 2000; Pintó *et al.* 2012).

HAV is the sole member of the genus *Hepatovirus* within the family *Picornaviridae* (Costa-Mattioli *et al.* 2002; Spradling *et al.* 2009; Knowles *et al.* 2012). The infectious viral particle of HAV consists of a non-enveloped icosahedral capsid that encloses a 7.5 kb positive-sense single-stranded RNA genome (Costa-Mattioli *et al.* 2003; Dotzauer 2008; Spradling *et al.* 2009; Knowles *et al.* 2012). HAV is antigenically stable with a single serotype (Nainan *et al.* 2006; Pintó *et al.* 2012). Infection with the virus confers lifelong immunity and an effective vaccine was developed to protect individuals that have not been previously exposed (Nainan *et al.* 2006; Pérez-Sautu *et al.* 2011a; Pintó *et al.* 2012).

However, recent publications have reported the detection of 'escape' mutants which could result in vaccination failure (Pérez-Sautu et al. 2011b; Pintó et al. 2012). The capsid gene VP1 and the junction VP1/P2B are used to genotype HAV strains (Nainan et al. 2006). The junction is used for rapid genotyping and the capsid gene for identification of possible vaccine escape mutants. Based on the nucleic acid sequence of the entire VP1 region six genotypes, with their subdivisions (A and B), have been defined: I, II and III grouping HAVs associated with humans, and IV, V and VI associated with non-human primates (Costa-Mattioli et al. 2002). These genotypes have a distinct geographical distribution: genotype IA prevails worldwide except in South Africa (SA) and the rest of the African continent where genotype IB predominates (Robertson et al. 1992; Taylor 1997).

The incidence of HAV depends on socio-economic standards like sanitary conditions or access to clean water sources. In much of the developing world, where such conditions are relatively poor, infection occurs early in childhood and almost all adults are immune to the disease. In industrialised countries, HAV has a low to moderate endemicity. As a result, most adults who have not been infected during childhood are susceptible to the infection, leading to the occurrence of hepatitis A outbreaks (Pintó et al. 2010). Hepatitis A is hyperendemic in SA with characteristics of both developing and industrialised countries (Robertson et al. 1992; Taylor 1997; Venter et al. 2007).

HAV is also one of the leading causes of viral foodborne infection (Bosch et al. 2011; Rodríguez-Lázaro et al. 2012). Food products can be contaminated directly or indirectly through the use of water contaminated with faeces containing HAV. This is mainly due to the fact that sewage treatment is unable to completely remove or inactivate viral particles which in turn may contaminate surface waters that can be used to irrigate or wash food products (Pintó et al. 2010). The faecal contamination of water in shellfish growing beds and irrigation water used for fresh produce has been implicated as the source of virus in foodborne outbreaks of hepatitis A. To date, the largest outbreak of hepatitis A recorded involved the consumption of sewage-contaminated clams and approximately 300,000 people became ill (Halliday et al. 1991; Nainan et al. 2006). The use of contaminated water to mix agrichemicals or inadequate sanitary facilities for food handlers at the point of harvest, at commercial outlets or at preparation facilities also poses risk of spreading the virus.

HAV has been detected in surface water sources in SA (Taylor et al. 2001) and on fresh produce at the point of retail (Netshikweta 2011). The risk of infection following ingestion of water for the different communities has also been quantified (Venter et al. 2007). Since the studies of Taylor (1997) there have been no molecular epidemiological investigations in SA to establish whether or not different genotypes or mutant strains are emerging or circulating in the country. Analysing water samples provides a passive surveillance system as to what HAV types are circulating in the environment as well as the surrounding communities. In addition, it provides valuable baseline data for molecular source tracking in the event of an outbreak. The aim of this study was therefore to characterise HAV strains detected in selected SA surface waters.

## METHODS

### Water samples

#### Surface water sample collection

From January 2012 to August 2012, 21 water samples (10 L) were collected monthly (except in February and July) from six sites near Polokwane, Limpopo (L) province, as part of the SA Water Research Commission (WRC) project K5/1875//4 (Figure 1). These samples originate from rivers and dams used to irrigate fresh produce on a large commercial farm. The samples were transported in cooler bags with
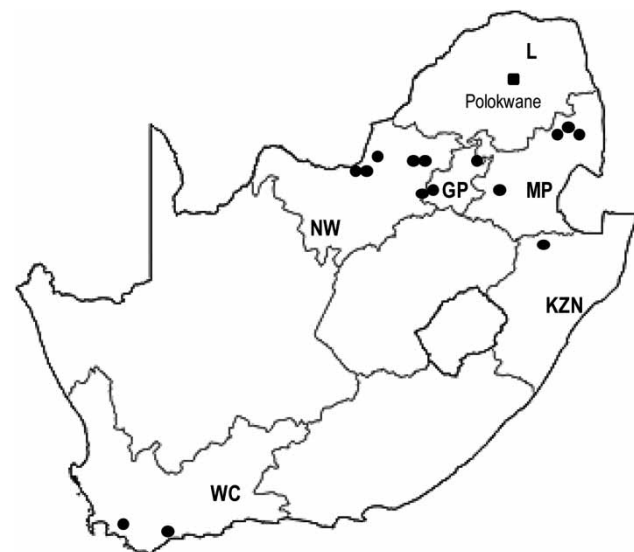


**Figure 1** | Distribution of surface water (■) and wastewater (•) sampling sites among six provinces of South Africa. GP, Gauteng province; KZN, Kwazulu-Natal province; L, Limpopo province; MP, Mpumalanga province; NW, North West province; WC, Western Cape province (adapted from Murray et al. 2013).

cold packs to the Department of Medical Virology, Faculty of Health Sciences, University of Pretoria. The temperature and pH of each sample were recorded upon arrival. The samples were either processed immediately or stored at 4 °C before being processed within 24 h of receipt. The sites from which samples were collected were: Bloed River (site 1) (two samples); River water drawn from below the Polokwane sewage works before the Sandrivier (site 5) (two samples); River on fresh produce farm (site 7) (five samples); Dam A at pump site on fresh produce farm (site 16B) (two samples); Dam B at pump site on fresh produce farm (site 17B) (five samples); River water from below the Seshego sewage works in the township (site 23) (two samples). Water was also collected from an irrigation pivot at the farm, at two points (three samples). Each sample was given a unique code comprising site number and the date at which it was taken.

## Water sampling at wastewater treatment works (WWTW)

From August 2010 to December 2011, a retrospective analysis was performed on 51 WWTW samples (500 mL) that were collected from the outflow of WWTW in three SA provinces: Gauteng (GP, nine samples), Mpumalanga (MP, 20 samples) and North West (NW, 16 samples) (Figure 1). In addition, sludge was collected in a treatment plant in Kwa-Zulu-Natal (KZN, two samples) and in the Western Cape (WC, two samples) (Figure 1), two samples were of unknown provincial origin. The 51 WWTW samples were referred to the Department of Medical Virology, Faculty of Health Sciences, University of Pretoria for microbial indicator analysis. These samples were selected because they had (thermotolerant) faecal coliform counts greater than $10^6$ colony forming units (cfu) per 100 mL (Murray *et al.* 2013).

## Microbial indicator organisms count

The *Escherichia coli* and thermotolerant (faecal) coliform counts, expressed as cfu/100 mL were determined for the surface water samples and the WWTW samples, respectively. The choice of method for the different matrices was specified in the protocols of each sample type. The membrane filtration technique coupled with m-ColiBlue24® broth (Millipore Corp., Billerica, MA) or M-FC agar (Standard Methods for the Examination of Water and Wastewater) was used to determine the *E. coli* and thermotolerant (faecal) coliform counts, respectively.

## Virus recovery and concentration

### Surface water samples

Prior to recovery, 100 μL of mengovirus ($5 \times 10^4$ copies/10 μL) was added to each sample as a process control to monitor the efficiency of viral recovery and downstream nucleic acid extraction. Viruses were recovered from the 10 L water samples using a glass wool adsorption-elution method based on the method of Vilaginès *et al.* (1993) as described by Mans *et al.* (2013). After elution, viruses are further concentrated to a volume of 10 mL in phosphate-buffered saline pH 7.4 (PBS; Sigma-Aldrich Co., St. Louis, MO, USA), using polyethylene glycol (Merck Schuchardt OHC, Hohenbrunn, Germany) and sodium chloride (Merck) ($PEG_{8000}$/NaCl) precipitation. The recovered virus concentrates were stored at −20 °C.

### Wastewater treatment works samples

Viruses were concentrated from 75 or 100 mL WWTW samples to a volume of 2 mL in PBS pH 7.4, using the $PEG_{8000}$/NaCl precipitation. No prior viral recovery was performed as the sample volumes were too small. Virus concentrates were also stored at −20 °C.

## Nucleic acid extraction

Total nucleic acid was extracted from 1 mL of virus concentrates using the automated platform of the MagNA Pure LC instrument (Roche Diagnostics GmbH, Mannheim, Germany) and the semi-automated platform of the Nucli-SENS® EasyMAG® instrument (BioMérieux, Marcy l'Etoile, France). The MagNA Pure LC Total Nucleic Acid isolation Kit (Large volume) was used for viral extraction on the MagNA Pure LC System (Roche Diagnostics). Extraction on the MagNA Pure LC System (Roche Diagnostics) was done for virus concentrates of the WWTW samples, while the NucliSENS® EasyMAG® instrument (BioMérieux) was used for the surface water samples. In either method, extracted nucleic acid was eluted in 100 μL of elution buffer and stored at −70 °C.

## Mengovirus detection

Nucleic acid extracted from surface water samples were tested for mengovirus using a one-step real-time reverse-transcriptase polymerase chain reaction (RT-PCR) assay. The detection was performed using mengo@ceeramTools™ Kit

(Ceeram s.a.s, La Chappelle-Sur-Erdre, France). The detection was performed on 5 µL of RNA extracts and the conditions of the assay were as follows: 45 °C for 10 min, 95 °C for 10 min and 45 cycles of 95 °C for 15 s, 60 °C for 45 s. The primers (Table 1) included in the kit, amplify a 99 bp fragment within the 5′noncoding region (5′NCR) of the mengovirus genome (Pintó et al. 2009).

## HAV detection

Nucleic acid (5 µL) from each sample (surface water and WWTW) was tested for HAV using a one-step real-time RT-PCR assay. Detection of HAV in surface water samples was performed using a hepatitisA@ceeramTools™ Detection Kit (Ceeram s.a.s) while detection in WWTW samples was performed using a RNA Ultrasense™ One-step qRT-PCR system kit (Invitrogen, Carlsbad, CA). Both assays were carried out in a LightCycler 2.0 (Roche Diagnostics) and used primers which amplified a 172 bp fragment within the most conserved region of the HAV genome, the 5′NCR (Table 1). The hepatitisA@ceeramTools™ Detection Kit (Ceeram s.a.s) includes an internal control to monitor the amplification process. The RNA Ultrasense™ One-step qRT-PCR system kit (Invitrogen) does not include an internal control, but a positive control, the cell culture adapted HAV strain HM175 43c was added to each run to monitor the amplification process. The conditions for the

assay were different for each kit. For the hepatitisA@ceeramTools™ Detection Kit (Ceeram s.a.s) the conditions of the assay were as follows: 50 °C for 1 h, for the reverse transcription reaction, followed by a hot start of 95 °C for 10 min and 45 cycles of 95 °C for 15 s for denaturation, 60 °C for 1 min for annealing and 70 °C for 1 min for extension (Costafreda et al. 2006). The conditions of the assay for the RNA Ultrasense™ One-step qRT-PCR system kit (Invitrogen) were as follows: 50 °C for 45 min, for the reverse transcription reaction, followed by a hot start of 95 °C for 15 min and 50 cycles of 95 °C for 15 s for denaturation, 60 °C for 1 min for annealing and 65 °C for 1 min for extension (Netshikweta 2011).

## HAV genotyping

### Complementary DNA (cDNA) synthesis

Using random hexamers primers (Roche Diagnostics), cDNA (20 µL) was synthesised from 10 µL of RNA extracts from samples that tested positive for HAV. The synthesis was performed using a Thermo Scientific RevertAid™ Premium Reverse Transcriptase kit (Thermo Scientific, Waltham, MA). The protocol used was a slight modification of that of the manufacturer's instructions, as 50 units of the reverse transcriptase was used instead of 200 units.

**Table 1** | Sequences of primers and probe used in this study

| PCR | Region | Primer/Probe name | Nucleotide Sequence (5′ → 3′) | Position[a] | Reference |
|---|---|---|---|---|---|
| Real time – Mengo | | Mengo110-F | GCG GGT CCT GCC GAA AGT | – | Pintó et al. 2009 |
| | | Mengo209-R | GAA GTA ACA TAT AGA CAG ACG CAC AC | – | Pintó et al. 2009 |
| | | Mengo147-probe | FAM-ATC ACA TTA CTG GCC GAA GC-TAMRA | – | Pintó et al. 2009 |
| Real time – HAV | 5′NCR | HAV68-F | TCA CCG CCG TTT GCG TAG | 68–85 | Costafreda et al. 2006 |
| | | HAV240-R | GGA GAG CCC TGG AAG AAA G | 240–222 | |
| | | HAV150-probe | FAM-TTA ATT CCT GCA GGT TCA GG-TAMRA | 150–169 | |
| First round PCR | VP1 | HAV1-F | GTT TTG CTC CTC TTT ATC ATG CTA TG | 2167–2192 | Costa-Mattioli et al. 2002 |
| | | HAV2-R | AGT CAC ACC TCT CCA GGA AAA CTT | 3308–3285 | |
| | VP1/P2B | 2870P-F | GAC AGA TTC TAC ATT TGG ATT GGT | 2870–2893 | Nainan et al. 2006 |
| | | 3381N-R | CCA TTT CAA GAG TCC ACA CAC T | 3381–3360 | |
| Second round PCR | VP1 | 2172P-F | GCT CCT CTT TAT CAT GCT ATG GAT | 2172–2195 | Costa Mattioli et al. 2002, Nainan et al. 2006 |
| | | 3125N-R | CCT GCA TTC TAT ATG ACT CT | 3125–3106 | |
| | VP1/P2B | 2896P-F | CTA TTC AGA TTG CAA ATA CAA T | 2,896–2,918 | Nainan et al. 2006 |
| | | 3289N-R | AAC TTC ATT ATT TCA TGC TCC T | 3,289–3,268 | |

[a]Positions are relative to the genome of the HAV strain HM175 (accession number M14707).

## Amplification of the VP1 region

The full length VP1 region (900 bp) was amplified from cDNA (5 or 1 μL) by conventional PCR, using primers published by Costa-Mattioli et al. (2002) (Table 1).

The total volume of the reaction mix was 25 μL and contained the following: 5×KAPATaq™ HotStart buffer (Kapa Biosystems, Cape Town, South Africa), 10 mM dNTPs, and 10 μM each of the forward and reverse primers respectively, 25 mM MgCl$_2$, and 1.25 units of KAPATaq™ HotStart DNA polymerase (Kapa Biosystems). The conditions under which the VP1 region was amplified were: pre-denaturation at 94 °C for 3 min, follow by 45 cycles of 94 °C for 30 s, 50 °C for 30 s, 72 °C for 1 min 30 s, and, final extension at 72 °C for 5 min. If no positive amplicons were obtained after the first round of PCR, a second PCR was performed and the VP1 gene was amplified using internal primers (Costa-Mattioli et al. 2002; Nainan et al. 2006) (Table 1) and 0.5 μL of the first PCR reaction as template. The assay conditions for the second round of PCR are essentially the same, except for the annealing temperature that decreased from 50 to 48 °C and the number of cycles that went from 45 to 35 cycles.

## Amplification of the VP1/P2B junction

The VP1/P2B junction (∼350 bp), was amplified from the same cDNA (5 μL) used to amplify the VP1 region using primers published by Nainan et al. (2006) (Table 1). The total volume of the reaction mix was 25 μL and contained the following: 5×KAPATaq™ HotStart buffer (Kapa Biosystems), 10 mM dNTPs, and 10 μM each of the forward and reverse primers respectively, 25 mM MgCl$_2$, and 1.25 units of KAPATaq™ HotStart DNA polymerase (Kapa Biosystems). The conditions under which the VP1/P2B junction was amplified were: pre-denaturation at 94 °C for 3 min, follow by 35 cycles of 94 °C for 30 s, 45 °C for 45 s, 72 °C for 1 min, and, final extension at 72 °C for 5 min. If no PCR products were visible after the first round of PCR, a nested PCR was performed using internal primers (Nainan et al. 2006) (Table 1) and 0.5 μL of the first PCR reaction as template. The conditions for the second round of VP1/P2B amplification are similar to the ones used for the first round PCR, except for the annealing temperature that increased from 45 to 48 °C.

## Sequencing and cloning

Products from the PCR were run through a 2% agarose gel by electrophoresis. The gel was stained with ethidium bromide for visualisation. Positive amplicons were purified directly or from the gel using the Zymogen DNA Clean & Concentrator-25™ Kit or the Zymoclean™ Gel DNA Recovery Kit (Zymo Research, Irvine, CA) respectively. The primers used to amplify the different regions were used to sequence positive amplicons directly after amplification, in both the forward and reverse direction, using the ABI Prism BigDye® Terminator v3.1 Cycle sequencing Kit on an ABI 3130 automated analyser (Applied Biosystems), as per manufacturer's instructions. In cases where more than one strain were detected in a sample, PCR products were cloned using the cloneJET™ PCR cloning kit (Fermentas, Canada). After colony PCR, at least 10 clones were randomly selected for each sample and sequenced using the pJet 1.2/blunt specific primers.

## Nucleotide sequence analyses

The sequences obtained were analysed and edited with Sequencher™ v4.10.1 and BioEdit v6.0.5. A BLAST search (http://0-blast.ncbi.nlm.nih.gov.innopac.up.ac.za/) was performed to verify the identity of the output sequences. Thereafter the edited sequences were compared to each other and to reference sequences retrieved from Genbank by pairwise comparison. The sequences obtained along with reference sequences for HAV and closely matched sequences from the output BLAST search were aligned in MAFFT version 6 (http://0-mafft.cbrc.jp.innopac.up.ac.za/alignment/server/index.html).

## Detection of novel or predicted mutation

After alignment, the sequences obtained were compared to reference sequences using the software BioEdit v6.0.5, to check for novel or predicted mutations at nucleotide and protein levels. Positions at which nucleotide mutations cause amino acid change potentially resulting in vaccine escape mutants (Pérez-Sautu et al. 2011b; Pintó et al. 2012) were carefully analysed.

## Phylogenetic analysis

Phylogenetic analyses were performed in MEGA 5 using the two-parameter model of Kimura from the neighbor-joining method. Two trees were generated (one for VP1 and one for VP1/P2B) and assessed by bootstrap analysis (1000 pseudoreplicates). The genotype of each sequenced strain was determined by how well it clusters with reference sequences on the phylogenetic tree.

## RESULTS AND DISCUSSION

### HAV detection

HAV was detected in 76% (16/21) of the surface water samples (Table 2). A number of the surface water samples were faecally contaminated as the *E. coli* counts were between 1 and 1,000 cfu/100 mL (data not shown). Mengovirus was detected in 43% (9/21) of the surface water samples where HAV was detected. The process control was detected in samples where no HAV was detected (site 16B and 17B on 2012-01-31) indicating that the processing and nucleic acid extraction procedures were adequate and the negative HAV result was therefore a true negative. The process control was not detected in samples where HAV was detected (site 5 on 2012-08-13; site 7 from 2012-05-16 to 2012-08-13; site 16B on 2012-08-13; site 17B on 2012-05-16; site 23 on 2012-08-13; site Pivit on 2012-05-16; site KPL1 on 2012-06-27) still indicating that the method for recovery and detection was appropriate. However, the process control was not detected in samples where HAV was not detected (site 1 on 2012-08-13; site 17B from 2012-06-27 to 2012-08-13)

(Table 2). The average recovery efficiency of mengovirus and HAV from turbid water using the glasswool adsorption-elution method is <1% and 5%, respectively (V. Ruhanya, unpublished data). The detection rate of HAV could therefore be an underestimation of the true prevalence of the virus in these surface waters.

HAV was detected in 37% (19/51) of the WWTW samples. A number of factors could account for a low detection rate in WWTW samples as opposed to surface water samples. No prior viral recovery was performed on the WWTW samples. Inhibitors, which were not completely removed during nucleic acid extraction, could have been co-concentrated during the $PEG_{8000}/NaCl$ precipitation step. As the assay used for the WWTW samples lacked an internal amplification control, the assay was repeated with diluted nucleic acid, to circumvent the possible effect of inhibitors. Noroviruses and sapoviruses were detected in the 19 WWTW samples that tested positive for HAV and in some of the remaining 31 samples (Murray *et al.* 2013). This indicates that the nucleic acid extraction was efficient and HAV was possibly present at levels below the detection limit of the assay.

**Table 2** | Summary of results of the analysis of surface water samples from the Polokwane sites for HAV

| Site no. | Date | Sample type | pH/ temp | Real-time RT-PCR analysis Mengovirus | HAV |
|---|---|---|---|---|---|
| 1 | 2012-01-31 | Bloed river | 7/22 | + | + |
| | 2012-08-13 | | 7.5/13.5 | − | − |
| 5 | 2012-01-31 | River water below the Polokwane sewage works | 1/24 | + | + |
| | 2012-08-13 | | 7.5/13.5 | − | + |
| 7 | 2012-01-31 | River water on fresh produce farm | 7/22 | + | + |
| | 2012-03-27 | | 7.5/16 | + | + |
| | 2012-05-16 | | 7.5/11 | − | + |
| | 2012-06-27 | | 7.5/11 | − | + |
| | 2012-08-13 | | 7.5/10.5 | − | + |
| 16B | 2012-01-31 | Dam A at pump site on fresh produce farm | 7/22 | + | − |
| | 2012-08-13 | | 7.5/12 | − | + |
| 17B | 2012-01-31 | Dam B at pump site on fresh produce farm | 7/22 | + | − |
| | 2012-03-27 | | 7.5/16 | + | + |
| | 2012-05-16 | | 7.5/11.5 | − | + |
| | 2012-06-27 | | 7.5/11.5 | − | − |
| | 2012-08-13 | | 8/12 | − | − |
| 23 | 2012-01-31 | River water below Seshego sewage works in the township | 7/23 | + | + |
| | 2012-08-13 | | 7.5/19 | − | + |
| Pivit | 2012-03-27 | Irrigation pivot | 7.5/16 | + | + |
| | 2012-05-16 | | 7.5/12 | − | + |
| KPL1 | 2012-06-27 | Irrigation pivot | 7.5/13 | − | + |

## Phylogenetic analysis

HAV was detected in 35 samples in total. Of the 35 samples, HAV strains could be genotyped from 32 samples: 15 surface water samples and 17 WWTW samples. The junction VP1/P2B could be amplified from 29 of the samples (13 surface water samples and 16 WWTW samples), while the VP1 region could only be amplified from 17 samples (11 surface water samples and six WWTW samples). Based on the phylogenetic analysis of both the VP1 genomic region and the VP1/P2B junction all the strains typed in this study clustered within genotype IB (Figure 2 and 3, respectively). In 1997, it was established that genotype IB predominated in SA (Taylor 1997). None of the strains sequenced in this study grouped within a genotype other than IB, suggesting that since 1997 genotype IB still predominates in SA. However, additional HAV strains will have to be characterised before a conclusive statement in this regard can be made.

The strains sequenced in this study form a unique SA cluster within genotype IB (Figure 2: South African cluster; Figure 3: South African cluster A). The clusters are supported by high bootstrap values (99% on Figure 2 and 97% on Figure 3). However, few strains grouped outside the unique cluster on Figure 3 (South African cluster B; sample 10/1010 indicated with ■) but were still within genotype IB. All the samples that grouped outside the unique cluster were either collected from the effluents of sewage treatment plants or from rivers into which sewage effluents are discharged.

When analysing samples from the environment, multiple strains are expected to be characterised from a single sample. During amplification and sequencing of the VP1/P2B junction, no more than one strain was detected in each sample except for one. On the other hand, multiple strains were detected in 15 out of the 17 samples from which the VP1 region could be amplified. Although HAV is antigenically stable, some regions of its genome are more variable than others and therefore more informative. It is mainly for that reason that genotyping of the different HAV strains using nucleotide sequence analyses of the full VP1 region, instead of other genomic regions commonly used, was proposed (Costa-Mattioli et al. 2002). Furthermore, some amino acids encoded from the VP1 region contribute to the immunodominant antigenic site targeted by most antibodies that are directed against HAV (Costa-Mattioli et al. 2003; Knowles et al. 2012) and these residues are prone to mutation (Pérez-Sautu et al. 2011b; Pintó et al. 2012). The predicted amino acid sequence of the strains characterised in this study showed great similarity (>98%

in amino acid identity) to that of the type strain HM175 (accession number M14707). However, minor amino acid changes were detected. These changes did not occur on the residues that contribute to the immunodominant antigenic site. In addition, the mutations detected among the strains were not consistent, except one: a R298 K (Position is relative to the genome of the HAV strain HM175). Even though the mutation occurred at the end of the protein sequence, it was consistent in all strains for which the VP1 region could be amplified. An arginine ($R$) to lysine ($K$) mutation in the amino acid sequence has little effect on the final conformation of the VP1 protein. Despite a small structural difference on the side chain, both amino acids are positively charged. Therefore the strains bearing this mutation will still be antigenically stable. Although no mutants were detected in this study, the presences of multiple strains of HAV that differ in the VP1 region suggest the importance of implementing a surveillance system in order to monitor the possible emergence of new variants. This would not only help in molecular source tracking in the event of an outbreak but in implementing vaccination policies if needs be.

In 2009, a study was published on the epidemiology of hepatitis A in Germany, with special emphasis on imported and non-imported HAV infections (Faber et al. 2009). The study detected and characterised HAV in tourists that got infected while visiting SA. These strains formed a distinct cluster when compared to strains detected in individuals who got infected either in Germany or in other regions of the world. In Figure 3, the strains isolated from the tourists were included (indicated with ∆ on the tree). Phylogenetic analysis clustered the German strains within SA cluster A. The high bootstrap value at the node leading to the cluster indicates close relationship with strains characterised in this study. This further confirms the distinct geographical distribution of HAV, a fact that can be exploited when investigating the source of an outbreak.

HAV strains were previously detected on tomatoes and lettuce at the point of retail in the Tshwane district (Netshikweta 2011). The detected strains (indicated with ▲ on Figure 3) also group within SA cluster A. The high bootstrap value at the node indicates that strains isolated from the water samples are similar to those isolated from the fresh produce. Furthermore, the strain isolated from the lettuce is identical to a strain detected in a WWTW sample (bootstrap value of 97%). The sample in question was collected from the Mpumalanga region (data not shown). Fresh produce can be contaminated at various points throughout the chain of production. Irrigation with surface water
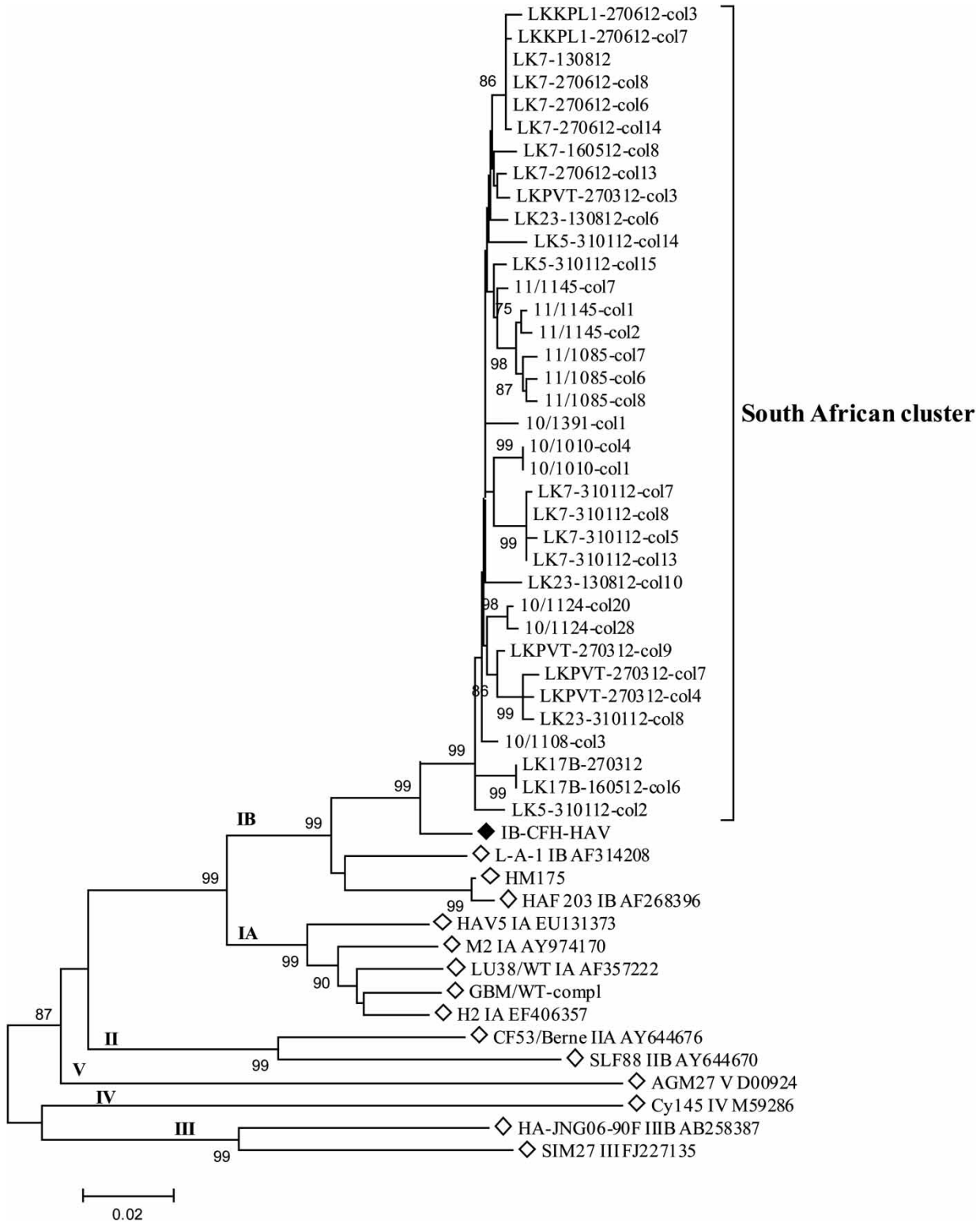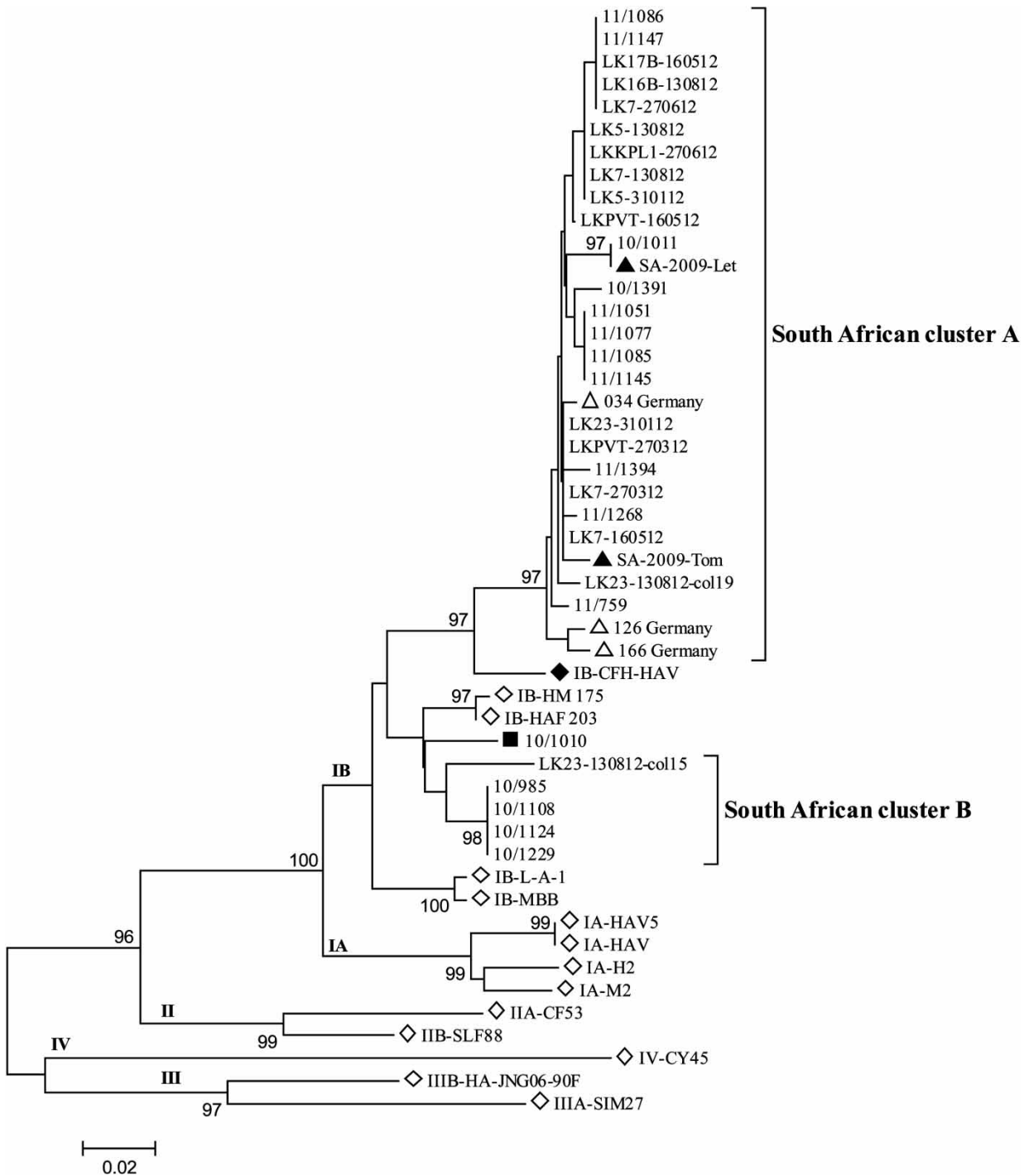
**Figure 2** | Phylogenetic tree based on the sequence of the full VP1 region (900 bp) of strains characterised in this study. The tree was constructed using the Kimura two-parameter model of the neighbour-joining method, with 1000 bootstrap replicates; bootstrap values ≥75% are indicated. Strains indicated with ◊ or ◆ are reference strains or strains from previous studies that were retrieved from Genbank. Strains with 'col' and a number at the end represent clones from strains characterised in this study.

**Figure 3** │ Pylogenetic tree based on the sequence of the VP1/P2B junction (∼ 350 bp) of strains characterised in this study. The tree was constructed using the Kimura two-parameter model of the neighbour-joining method, with 1000 bootstrap replicates; bootstrap values ≥75% are indicated. Strains indicated with ◊ or ♦, and, ▲ or △ are reference strains or strains from previous studies that were retrieved from Genbank. Strains with 'col' and a number at the end represent clones from strains characterised in this study.

contaminated with faeces containing HAV and agrichemicals mixed with contaminated water are possible sources of contamination. The lack of proper washing facilities for food handlers on the farm, at commercial outlets or during preparation are additional sources of contamination. In the present case the exact source of contamination of the produce could not be defined, but the similarity with a strain detected in sewage effluent suggests that sewage

treatment in the region of the WWTW sample does not remove all HAV particles.

## CONCLUSIONS

HAV strains of genotype IB are still prevalent in SA as no other genotypes were detected in the water samples. The presence of genotype IB in the surface water sources used to irrigate fresh produce indicates human faecal contamination. Therefore, faecally-contaminated water sources are a potential vehicle of transmission of HAV infection and a potential source of contamination of irrigated fresh produce in SA. However, it would be essential to characterise HAV from infected individuals in order to assess whether or not the strains detected in the environment are clinically relevant.

## ACKNOWLEDGEMENTS

## REFERENCES

Bower, W. A., Nainan, O. V., Han, X. & Margolis, H. S. 2000 Duration of viremia in hepatitis A virus infection. *The Journal of Infectious Diseases* **182**, 12–17.

Collier, L. & Oxford, J. 2006 Introduction to the hepatitis viruses. In: *Human Virology. A Text for Students of Medicine, Dentistry, and Microbiology* (L. Collier, J. Oxford & J. Pipkin, eds). Oxford University Press Inc, New York, pp. 159–160.

Costafreda, M. I., Bosch, A. & Pintó, R. M. 2006 Development, evaluation, and standardization of a real-time Taqman reverse transcription-PCR assay for the quantification of hepatitis A virus in clinical and shellfish samples. *Applied and Environmental Microbiology* **72**, 3846–3855.

Costa-Mattioli, M., Cristina, J., Romero, H., Perez-Bercof, R., Casane, D., Colina, R., Garcia, L., Vega, I., Glikman, G., Romanowsky, V., Castello, A., Nicand, E., Gassin, M., Billaudel, S. & Ferré, V. 2002 Molecular evolution of hepatitis A virus: a new classification based on the complete VP1 protein. *Journal of Virology* **76** (18), 9516–9525.

Costa-Mattioli, M., Di Napoli, A., Ferré, V., Billaudel, S., Perez-Bercoff, R. & Cristina, J. 2003 Genetic variability of hepatitis A virus. *The Journal of General Virology* **84**, 3191–3201.

Dotzauer, A. 2008 Hepatitis A virus. In: *Encyclopedia of Virology* (B. W. J. Mahy & M. V. van Regenmortel, eds). 3rd edn., Academic Press, Slovenia, pp. 343–350.

Faber, M. S., Stark, K., Behnke, S. C., Schreier, E. & Frank, C. 2009 Epidemiology of hepatitis A infection, Germany, 2007–2008. *Emerging Infectious Diseases* **15**, 1760–1768.

Halliday, M. L., Kang, L. Y., Zhou, T. K., Hu, M. D., Pan, Q. C., Fu, T. Y., Huang, Y. S. & Hu, S. L. 1991 An epidemic of hepatitis A attributable to the ingestion of raw clams in Shanghai, China. *The Journal of Infectious Diseases* **164**, 852–859.

Knowles, N. J., Hovi, T., Hyypiä, T., King, A. M. Q., Lindberg, A. M., Pallansch, M. A., Palmenberg, A. C., Simmonds, P., Skern, T., Stanway, G., Yamashita, T. & Zell, R. 2012 Family Picornaviridae. In: *Virus Taxonomy. Ninth Report of the International Committee on Taxonomy of Viruses* (A. M. Q. King, M. J. Adams, E. B. Carstens & E. J. Lefkowitz, eds). Elsevier Academic Press, San Diego, pp. 855–880.

Mans, J., Netshikweta, R., Magwalivha, M., Van Zyl, W. B. & Taylor, M. B. 2013 Diverse norovirus genotypes identified in sewage-polluted river water in South Africa. *Epidemiology and Infection* **141**, 303–313.

Murray, T. Y., Mans, J. & Taylor, M. B. 2013 Human calicivirus diversity in wastewater in South Africa. *Journal of Applied Microbiology* **114**, 1843–1853.

Nainan, O., Xia, G., Vaughan, G. & Margolis, H. S. 2006 Diagnosis of hepatitis A virus infection: a molecular approach. *Clinical Microbiology Reviews* **19**, 63–79.

Netshikweta, R. 2011 *Optimisation and Assessment of Real-time PCR Techniques for the Detection of Selected Food- and Waterborne Viruses*. MSc dissertation, University of Pretoria, Pretoria, viewed 130428 < http://upetd.up.ac.za/thesis/available/etd-05222012-143527.

Pérez-Sautu, U., Costafreda, M. I., Lite, J., Sala, R., Barrabeig, I., Bosch, A. & Pintó, R. M. 2011a Molecular epidemiology of hepatitis A virus infections in Catalonia, Spain, 2005–2009: circulation of newly emerging strains. *Journal of Clinical Virology* **52**, 98–102.

Pérez-Sautu, U., Costafreda, M. I., Caylà, J., Tortajada, C., Lite, J., Bosch, A. & Pintó, R. M. 2011b Hepatitis A virus vaccine escape variants and potential new serotype emergence. *Emerging Infectious Diseases* **17** (4), 734–737.

Pintó, R. M., Costafreda, I. & Bosch, A. 2009 Risk assessment in shellfish-borne outbreaks of hepatitis A. *Applied and Environmental Microbiology* **75**, 7350–7355.

Pintó, R. M., Costafreda, I., Pérez-Rodriguez, F. J., D'Andrea, L. & Bosch, A. 2010 Hepatitis A virus: State of the art. *Food and Environmental Virology* **2**, 127–135.

Pintó, R. M., D'Andrea, L., Pérez-Rodriguez, F. J., Costafreda, M. I., Ribes, E., Guix, S. & Bosch, A. 2012 Hepatitis A virus evolution and the potential emergence of new variants escaping the presently available vaccines. *Future Microbiology* **7**, 1–15.

Robertson, B. H., Jansen, R. W., Khanna, B., Totsuka, A., Nainan, O. V., Siegl, G., Widell, A., Margolis, H. S., Isomura, S., Ito, K., Ishizu, T., Moritsugu, Y. & Lemon, S. M. 1992 Genetic relatedness of hepatitis A virus strains recovered from different geographical regions. *The Journal of General Virology* **73**, 1365–1377.

Rodríguez-Lázaro, D., Cook, N., Ruggeri, F. M., Sellwood, J., Nasser, A., Nascimento, M. S. J., D'Agostino, M., Santos, R., Saiz, J. C., Rzeżutka, A., Bosch, A., Gironés, R., Carducci, A., Muscillo, M., Kovač, K., Diez-Valcarce, M., Vantarakis, A., von Bonsdorff, C.-H., Husman, A. M. R., Hernández, M. & van der Poel, W. H. M. 2012 Virus hazards from food, water and other contaminated environments. *FEMS Microbiology Reviews* **36**, 786–814.

Spradling, P. R., Martin, A. & Feinstone, S. 2009 Hepatitis A Virus. In: *Clinical Virology* (D. D. Richman, R. J. Whitley & F. G. Hayden, eds). ASM Press, Washington DC, pp. 1083–1108.

Standard Methods for the Examination of Water and Wastewater 2005 21st edn., American Public Health Association/ American Water Works Association/Water Environment Federation, Washington DC, USA

Taylor, M. B. 1997 Molecular epidemiology of South African strains of hepatitis A virus:1982–1996. *Journal of Medical Virology* **51**, 273–279.

Taylor, M. B., Cox, N., Very, M. A. & Grabow, W. O. K. 2001 The occurrence of hepatitis A and astroviruses in selected river and dam waters in South Africa. *Water Research* **35**, 2653–2660.

Venter, J. M., van Heerden, J., Vivier, J. C., Grabow, W. O. K. & Taylor, M. B. 2007 Hepatitis A virus in surface water in South Africa: what are the risks? *Journal of Water and Health* **5**, 229–240.

Vilaginès, P., Sarrette, B., Husson, G. & Vilaginès, R. 1993 Glass wool for virus concentration at ambient water pH level. *Water Science and Technology* **27**, 299–306.