Reviews in Medical Virology

Rev. Med. Virol. 2006; **16**: 365–383.
Published online 23 August 2006 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/rmv.515

REVIEW

Virus discovery by sequence-independent genome amplification

Helen E. Ambrose and Jonathan P. Clewley*

Virus Reference Department, Centre for Infections, Health Protection Agency, London NW9 5EQ, UK

SUMMARY

Genome sequences from several blood borne and respiratory viruses have recently been recovered directly from clinical specimens by variants of a technique known as sequence-independent single primer amplification. This and related methods are increasingly being used to search for the causes of diseases of presumed infectious aetiology, but for which no agent has yet been found. Other methods that do not require prior knowledge of the genome sequence of any virus that may be present in the patient specimen include whole genome amplification, random PCR and subtractive hybridisation and differential display. This review considers the development and application of these techniques. Copyright © 2006 John Wiley & Sons, Ltd.

Received: 31 May 2006; Revised: 2 June 2006; Accepted: 5 June 2006

INTRODUCTION

Since the early 1980s, viral genomes have been recovered by molecular biological methods from clinical specimens and their sequences have subsequently been deduced (Table 1). Often this has been the first step in the characterisation of previously unknown viruses, and has led to the expression of viral proteins for use in diagnostic assays (e.g. EIAs) and even to growth of the virus in cell culture from molecular clones (e.g. HCV). The genomes of the viruses that have been discovered by molecular methods are diverse, and encompass almost the full range of RNAs and DNAs, single or double stranded, segmented, linear or circular (Table 1). The first methods were devised before the invention of PCR. They ranged from simple cloning (for parvovirus B19) to recombinant cDNA library construction (for HCV) [1–3]. With the advent of PCR, primers based on conserved regions of viral genomes (e.g. the polymerase gene) were designed to amplify sequences from novel viruses for example retroviruses [4–6]. These degenerate or consensus primers are generally only useful when searching for a specific sort of virus genome.

Several related PCR methods were developed in the 1990s involving, first, ligation of primer binding sites (known as adapter or linker oligonucleotides) to DNA fragments and, second, sequence enrichment by amplification. Two of the more well known methods are representational difference analysis (RDA) and sequence-independent single primer amplification (SISPA). Although RDA has been successfully used to discover at least three viruses, SISPA-type methods have been more popular in the last few years (Table 1). This review will discuss these methods and others, and will attempt to show where they may be usefully applied. A previous review of this subject was by Muerhoff and colleagues [7].

SISPA

SISPA was introduced by Reyes *et al* [8] as a technique to identify viral nucleic acid of unknown sequence present at low concentration. It was used to amplify cDNA prepared from 10 μg of nucleic acid extracted from 1.5 g of a faecal sample

*Corresponding author: J. P. Clewley, Virus Reference Department, Centre for Infections, Health Protection Agency, London NW9 5EQ, UK. E-mail: jonathan.clewley@hpa.org.uk

Abbreviations used

AFLP, amplified fragment length polymorphism; aRNA, antisense RNA; BLAST, basic local alignment search tool; CoV, coronavirus; DD, differential display; DOP-PCR, degenerate oligonucleotide primed PCR; MDA, multiple displacement amplification; PEP-PCR, primer extension preamplification PCR; PHACCS, phage communities from contig spectrum; RDA, representational difference analysis; rPCR, random PCR; RT-PCR, reverse transcriptase PCR; SARS, severe acute respiratory syndrome; SARS-CoV, severe acute respiratory syndrome coronavirus; SISPA, sequence-independent single primer amplification; TTV, torque teno virus; VIDISCA, virus-discovery-cDNA-AFLP; WGA, whole genome amplification.

Table 1. Some examples of viral genomes initially discovered or characterised by molecular methods

Virus	Disease	Advance knowledge*	Genome nucleic acid	Methods used	Year	Reference
Parvovirus B19	Fifth disease	Yes	ssDNA	Molecular cloning and DNA hybridization	1984	[1,2]
Hepatitis C virus	Non-A, non-B hepatitis	No	ssRNA	Transmission in primates; molecular cloning and immunoscreening	1989	[3]
Hepatitis E virus	Non-A, non-B hepatitis	No	ssRNA	Transmission in primates; molecular cloning and sequence similarity	1990	[75,76]
Retroviruses	Various	Yes	ssRNA and proviral dsDNA	Degenerate or consensus primer PCR and sequencing	1990 onwards	[4–6]
Rotavirus	Gastroenteritis	Yes	dsRNA segmented	SISPA	1992	[14]
Astrovirus	Gastroenteritis	Yes	ssRNA	SISPA and immunoscreening	1993	[10]
HHV-8, KSHV	Kaposi's sarcoma, B cell lymphomas	No	dsDNA	Sequence enrichment by representational difference analysis (RDA)	1994	[77]
GB viruses A and B	Not known	No	ssRNA	Transmission in primates and RDA	1995	[46]
GB virus C	Not known	Yes	ssRNA	SISPA, cloning and sequencing	1996	[11]
TTV or Torque teno virus	Not known	No	ssDNA circular	RDA	1997	[47]
Bovine parvovirus	Not known	No	ssDNA	SISPA, cloning and sequencing	2001	[19]
Coronavirus, HCoV-NL63	Respiratory disease	No	ssRNA	Cell culture and VIDISCA	2004	[23]
Bocavirus	Respiratory disease?	No	ssDNA	Random PCR, cloning and sequencing	2005	[22]
Parvovirus 4 and TT-like viruses	Not known	No	ssDNA	SISPA, cloning and sequencing	2005	[25]

*Were the investigators searching for a specific type of virus genome?

DOI: 10.1002/rmv

associated with Norwalk virus gastroenteritis [9]. Cloning and immunoscreening of the SISPA-recovered DNA led to the first sequence of a norovirus genome, and also that of an astrovirus [10]. It was later used for the recovery of a 'hepatitis' G virus genome [11].

As initially described, the SISPA procedure was similar to previously published methods for cloning cellular mRNA and chromosomal DNA [12,13]. One method (Figure 1), described by Akowitz and Manuelidis [12], was referred to as primer-directed enzymatic amplification, and was developed to make cDNA libraries from small amounts of mRNA. DNA linker-adapters were generated by annealing together complementary 12-mer and 19-mer oligonucleotides and then ligating them to the blunt ends of cDNA. The 12-mer was phosphorylated at the 5′ end to increase the efficiency of ligation because by doing this both DNA strands could be covalently joined together by a phosphodiester bond. As the 19-mer oligonucleotide was complementary to both ends of the ligated cDNA, it could be used as a primer for PCR amplification. In this way, a representative bacteriophage λgt10 cDNA library was created from about 20 pg of globin mRNA. Another method (Figure 2), described by Johnson [13],

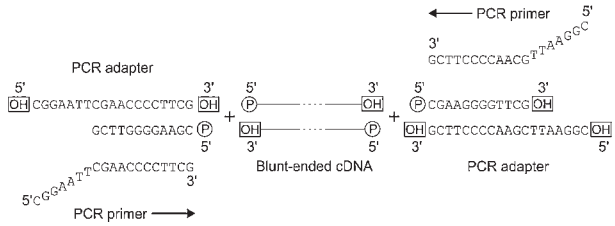

Figure 1. A linker-adapter with an overhang and one phosphorylated 5′ end. The synthetic oligonucleotides used in the work of Akowitz and Manuelidis [12]. The 12-mer oligonucleotide was exposed to a kinase to add a 5′-phosphate group, and was then annealed with the 19-mer oligonucleotide to make the adapter-linker. This adapter was then annealed to the blunt-ended cDNA. The cDNA contains the region of interest whose sequence is unknown. The 19-mer oligonucleotide was used as a PCR primer to amplify the cDNA, as it was able to anneal to both ends of the cDNA with the attached adapters-linkers. The 5′-phosphate group on the 12-mer oligonucleotide is intended to increase the efficiency of the ligation of the linker-adapter to the cDNA because by doing this, both DNA strands could be covalently joined together by a phosphodiester bond

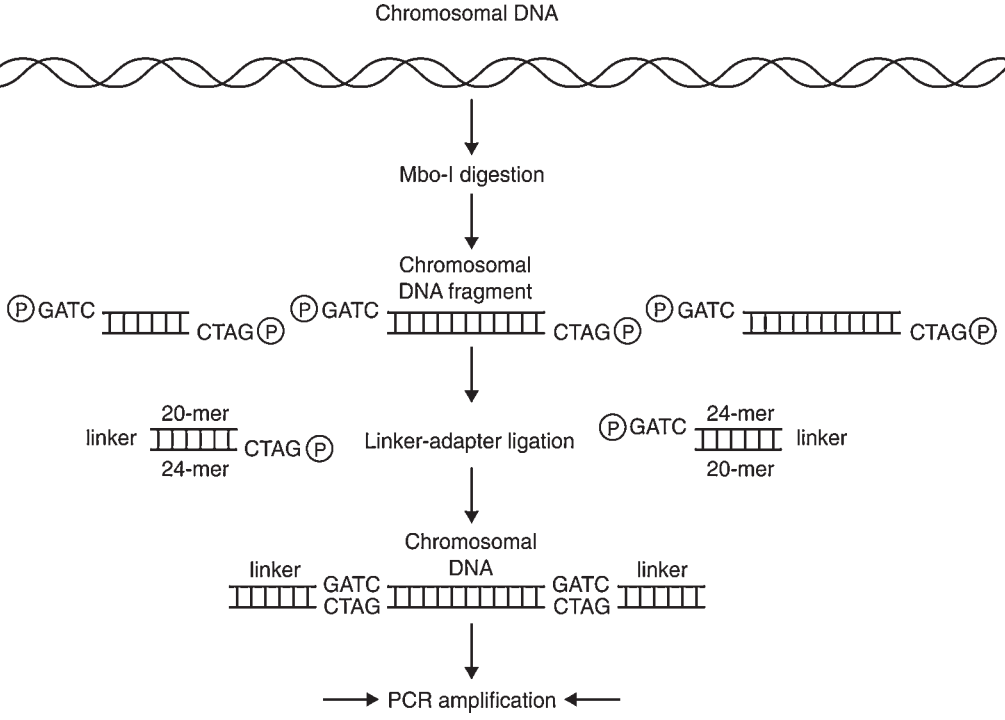

Figure 2. Ligation of a linker-adapter with an overhanging restriction site to DNA digested with the same restriction enzyme. The strategy of Johnson (1990) [13] for amplifying fragments of chromosomal DNA digested with *MboI* after ligation of *MboI* linker-adapters to the DNA. Initially chromosomal DNA of unknown sequence was digested with the restriction enzyme *MboI*. The linker adapters were made from 20-mer and 24-mer oligonucleotides, with the 5′ end of the 24-mer oligonucleotide exposed to a kinase to add a phosphate group, for more efficient ligation. These linker-adaptors were ligated onto the *MboI* digested chromosomal DNA. The 20-mer oligonucleotide was subsequently used as a primer to PCR amplify the chromosomal DNA of unknown sequence

DOI: 10.1002/rmv

involved the digestion of chromosomal DNA by the restriction enzyme *Mbo*I. A linker-adapter with an overhanging *Mbo*I site was made from a complementary pair of oligonucleotides (a 5′-phosphorylatyed 24-mer and a complementary 20-mer). This adapter was then ligated to the *Mbo*I digested chromosomal DNA. The 20-mer oligonucleotide was subsequently used as a PCR primer.

The original formulation of SISPA (Figure 3) used elements of both these methods [8]. It involved the directional ligation of an asymmetric adapter (referred to as a linker/primer in the original paper) onto both termini of blunt-ended cDNA. The common end sequence of the adapter allowed the cDNA to be amplified in a subsequent PCR using a single primer. Restriction enzyme sites (*Eco*RI and *Not*I) were included within the linker, so that the DNA generated from the PCR amplification could be cloned into a vector using these sites and subsequently easily sequenced. In addition, each primer had half an *Nru*I site so that if the adapters ligated together to form dimers, they could be removed by digestion with the enzyme *Nru*I. The *Nru*I digest would leave

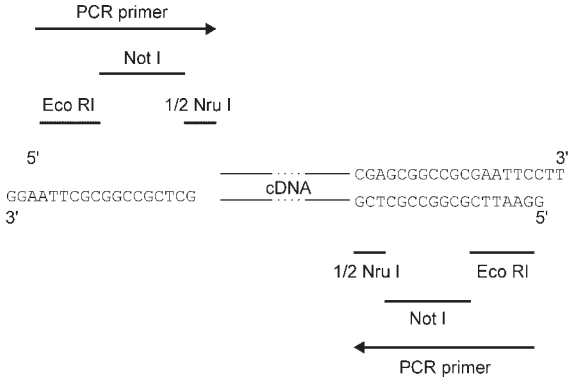

Figure 3. The adapters (linker/primers) used in the original SISPA publication of Reyes and Kim. The adapters were made by annealing together the top and bottom synthetic oligonucleotides. They were ligated to cDNA of unknown sequence. The overlap at one end of the hybrid ensured that the other, blunt end of the adapters ligated with the blunt ends of the cDNA. The adapters contained *Eco*RI and *Not*I restriction sites so that amplified DNA could be inserted into cloning vectors efficiently due to the sequence and size of these restriction sites. *Eco*RI and *Not*I are enzymes that cut DNA infrequently and are unlikely to occur very often in the cDNA. Hence, they are sites commonly engineered into cloning vectors. The adapters contain half an *Nru*I site at their blunt end. Therefore, if two adapters ligate together to form a dimer an *Nru*I site is formed. Digestion with *Nru*I prior to the ligation will cleave the dimer back into two monomers. This step is necessary because the dimers may preferentially ligate into the cloning vector and so reduce the efficiency of the cDNA cloning (Reyes *et al*. [8])

the blunt-end termini of the adapters with a phosphate group. There was no need, therefore, for the oligonucleotides to have phosphate groups added to them with the enzyme polynucleotide kinase, as was used in the procedures previously described.

SISPA for rotaviruses

A year later, Lambden and Clarke developed a SISPA methodology for dsRNA viruses (Figure 4) [14,15]. They demonstrated the feasibility of their method using a human rotavirus group C genome segment of 728 base pairs. The oligonucleotide ligated onto the dsRNA segment was blocked at the 3′ end with an amino group and phosphorylated at the 5′ end. Once the RNA strands had been separated, the ssRNA was reverse transcribed into cDNA. This was then amplified by PCR with a complementary primer. This method, and variations of it, has subsequently been used for recovery of various rotavirus genomes from clinical samples [16–18].

Enrichment of the target before SISPA

Over recent years, SISPA has been further developed to amplify both single and double-stranded RNA and DNA of heterogeneous size and sequence. Figure 5 gives an overview of the technique. Initially, there needs to be some pre-treatment of the sample to remove any non-viral nucleic acid, which might come from both bacteria and host nuclei or mitochondria. The methods used to separate viral from non-viral nucleic acid are based on the relatively small size of viral particles and the protection of the viral genome by the capsid and, if present, the envelope. For example, for purification of rotavirus dsRNA from faeces, Lambden *et al*. [14,15] treated the faecal suspension with ribonuclease T1 prior to extraction with the guanidinium-based reagent RNAzol B and silica particles (Geneclean II). The high concentrations of rotavirus present in faeces (e.g. 10^{10} particles) facilitated the extraction, and up to 100 ng dsRNA were recovered. Other viruses in other tissues may not yield as much genomic nucleic acid, and so the removal of extraneous DNA and RNA becomes even more critical. For the extraction of viral nucleic acid from blood, Allander *et al*. purified serum through 0.22 μm filters (to remove cells and mitochondria) and removed any free DNA by digestion with DNase I [19,20]. For viruses that are larger than 0.2 μm, for example herpesviruses, a caesium chloride gradient (density of

DOI: 10.1002/rmv

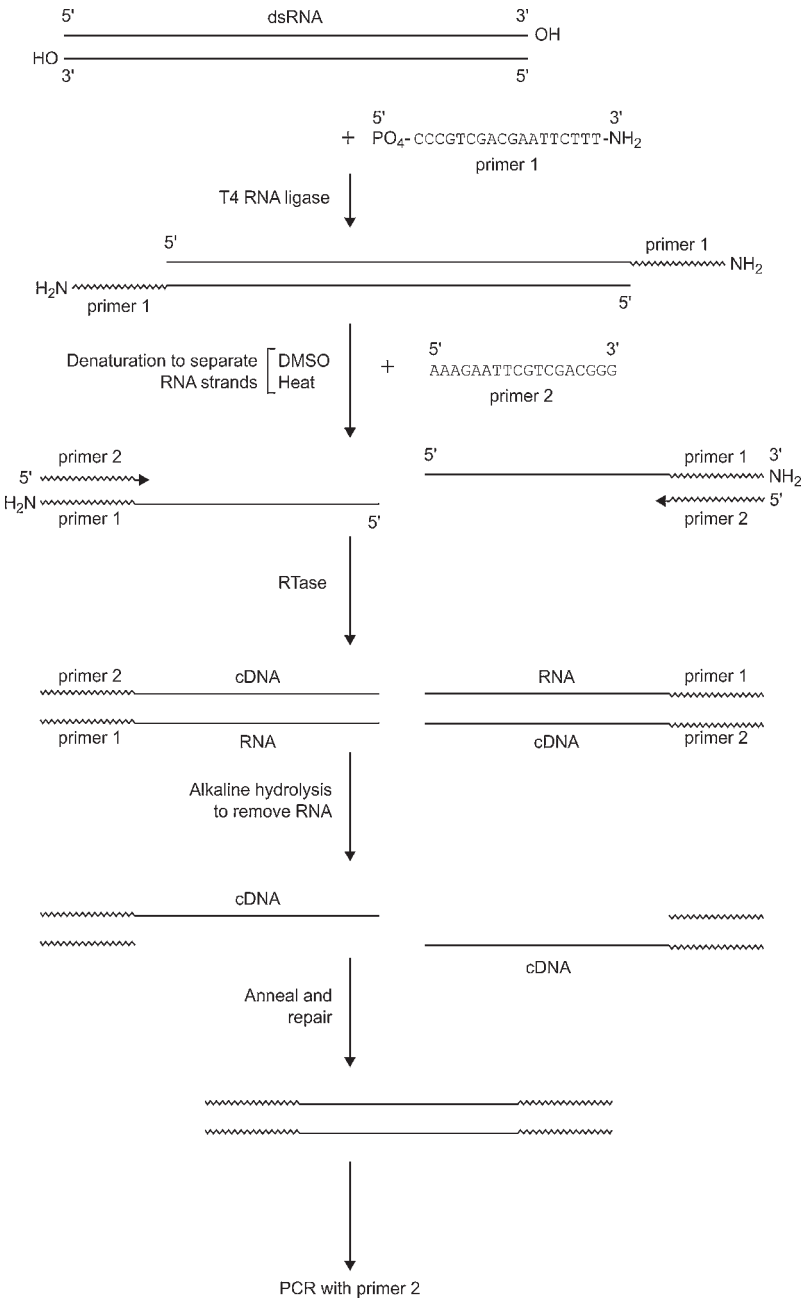

Figure 4. The dsRNA single primer amplification method of Lambden and colleagues. A single oligonucleotide (primer 1) was synthesized with a 3′-terminal amino group to prevent its concatenation during the following ligation reactions. The primer was also phosphorylated with ATP at the 5′-terminus to increase ligation efficiency as discussed in the main text, and the ATP was radiolabelled so that subsequent steps in the method could be monitored. This amino blocked and phosphorylated oligonucleotide was covalently attached using T4 RNA ligase to each 3′ end of the dsRNA segment. The dsRNA was denatured into two separate strands by heating in DMSO and a second oligonucleotide, complementary to the first, was then annealed to the primer-tailed ssRNA. Two single-stranded (ss) molecules with ds tails were thereby generated, and these served as templates for cDNA synthesis with reverse transcriptase. The RNA strand of the cDNA/RNA duplex was removed by hydrolysis with alkali, and the remaining cDNA strands were annealed together and repaired. The unknown cDNA sequence was then amplified by PCR using the second primer. To sequence the PCR products, they were cloned into the vector M13 mp*, which had been cleaved with *Sma*I and dephosphorylated (Lambden *et al.* [14])

DOI: 10.1002/rmv

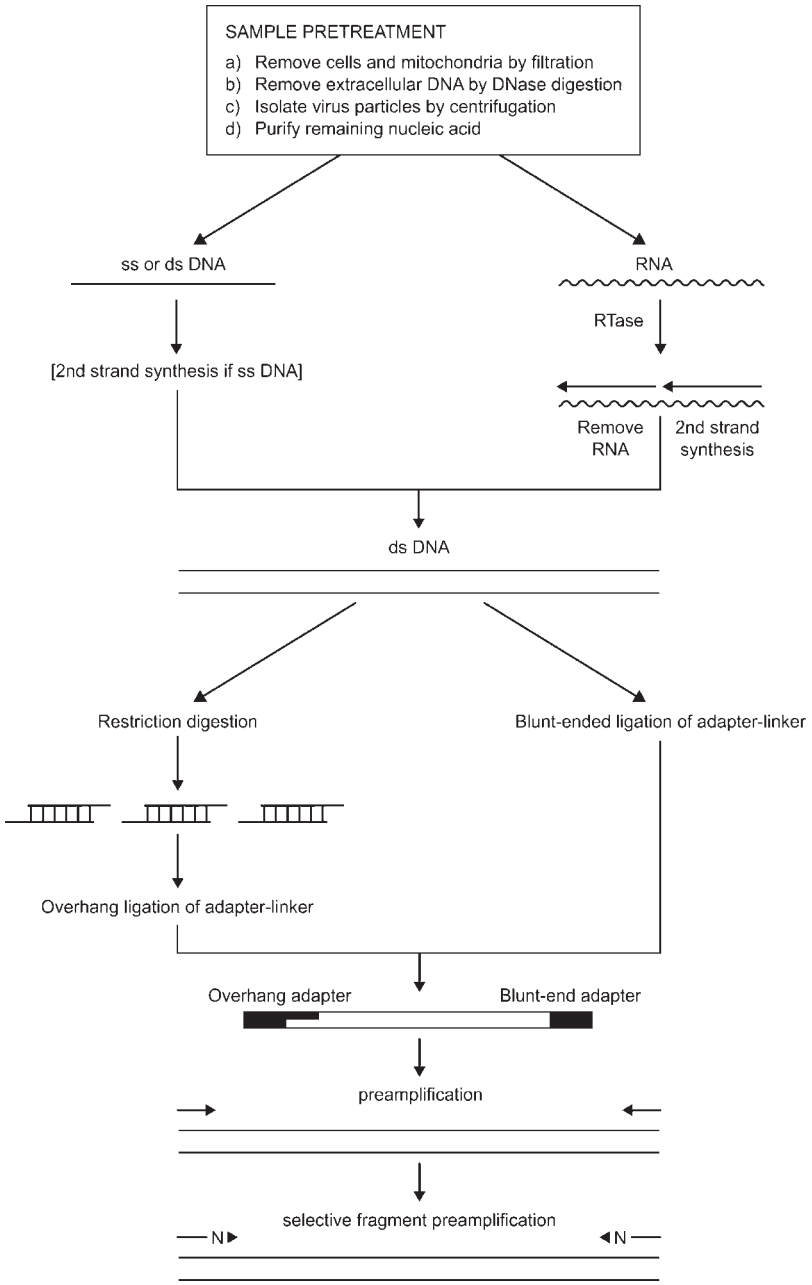

Figure 5. An overview of SISPA. This method is used to amplify viral nucleic acid of unknown sequence. Initially, the sample is pretreated by a combination of methods to aid purification of the viral nucleic acid. These methods include filtration of the sample to remove host cells and mitochondria, removal of extracellular DNA by DNase treatment and the isolation of viral particles by centrifugation. Finally, the remaining viral nucleic acid can be further purified by silica particle and guanidinium extraction methods. The sample may need to be divided at this stage if it is not known whether the viral nucleic acid is either RNA or DNA. If ssDNA is present, a second strand of DNA needs to be synthesized; if the sample contains RNA, cDNA needs to be generated. Once dsDNA has been produced, it can be digested by restriction enzymes. This enables the ligation of adapter-linkers with the relevant overhangs. Alternatively, adapter-linkers with blunt ends can be ligated onto the termini of undigested dsDNA. To amplify the unknown viral nucleic acid, a primer can be used that is complementary to the known sequence of the adapter-linker. A further selective round of amplification can be performed using a primer with an additional nucleotide at the 3′ end. By doing this, in theory, only a quarter of the DNA fragments will be amplified. Finally, the sequence of the amplified viral nucleic acid can be characterised by downstream cloning and sequencing. More detail on these methods can be found in the main text

DOI: 10.1002/rmv

virus particles usually exceeds 1.11 g/ml) or a sucrose cushion (usually 30%) could give a higher yield than filtration [21]. Another option prior to extraction is to concentrate viral particles by ultracentrifugation [22]. In some cases, it may be possible to use cell culture to amplify any virus present. If the cells show a cytopathic effect, nucleic acid can be extracted from virus purified from the culture supernatant [23].

The viral nucleic acid is then extracted with either commercial kits (e.g. Qiagen, Crawley, UK) or chaotropic agents, such as TRIzol (Invitrogen, Paisley, UK). Sometimes, it is necessary to divide the sample at this stage if there is a possibility that it contains both unknown RNA and DNA viruses [20,22]. It is necessary to generate dsDNA from the viral nucleic acid because it is the starting molecule for SISPA. For RNA viruses, cDNA is generated via a reverse transcriptase reaction [22]; alternatively reverse transcription may be incorporated as part of the SISPA reaction [14,24]. If an ssDNA virus is suspected, for example some parvoviruses which package only negative-stranded DNA, or the circoviruses with ssDNA circular genomes, then a second-strand DNA synthesis reaction needs to be incorporated.

Restriction digestion and adapter ligation

The majority of SISPA methods fragment the dsDNA with a restriction enzyme before the adapters are ligated [23,25]. As viral genomes are of limited genomic complexity, restriction enzymes with four base pair recognition sites are used, for example, *Csp 6.1* [25] or *Mse*I and *HinP1*I [23]. However, for some viral genomes for example HIV-1, *Eco*RI may be more appropriate than *HinP1*I [23]. The adapters are designed to have minimal self-complementary sequences, 58% GC bias for specificity in annealing reactions, and may contain restriction enzyme sites for downstream cloning [12]. The adapters can either be designed for blunt-ended ligation [8] or have overhangs complementary to the restriction enzyme used to fragment the viral DNA [23]. To prevent concatamerisation of the adapters, they can be amino-linked or phosphorylated at one end, as mentioned above [14].

After the ligation of the adapters, the dsDNA of unknown sequence can be amplified by PCR using a primer complementary to the known adapter sequence. Due to the low complexity of viral genomes, a restriction digest should produce a large amount of a limited number of fragments. After the amplification, these will be visible as discrete bands on an agarose gel [19]. Larger human or bacterial genomes are more complex than viral genomes; therefore many differently sized fragments are generated from a restriction digest. Amplification of these fragments would result in a DNA smear on an agarose gel. Ultimately, the resulting viral PCR amplicons can either be cloned and sequenced [19], immunoscreened [9,10], characterised by hybridisation reactions [8] or be identified via a DNA microarray [26–28]

The main advantages of SISPA are: (i) that it can be used to identify an unknown viral nucleic acid present in relatively limited amounts (10^6 copies) and, (ii) that it is culture independent [8,19,29]. The method can be applied to all kinds of clinical sample and can be used to identify both ss and dsRNA and DNA viruses [8,15,21]. The abundance of host and mitochondrial DNA in samples is the main disadvantage, and also the presence of other contaminating sequences (e.g. bacterial ones). To monitor for this, the extracted nucleic acid can be screened by PCR for human 18S rDNA and human mitochondrial DNA, and the results of both these PCRs would be expected to be negative [21]. Some clinical samples can have a very limited viral content so, ideally, samples should be taken during the viraemic phase when viral concentrations are highest.

The sensitivity of SISPA is dependent on both the characteristics of the clinical sample and the properties of the virus in question. The amount of nucleic acid that has been used for SISPA experiments has varied greatly. For example, after 30 cycles of amplification, 10 fg of ΦX174 bacteriophage generated sufficient amplicons to be visible on an agarose gel [8]. In comparison, sufficient full-length cDNA for SISPA was generated from 10 ng of each dsRNA group C rotavirus genome segment, extracted from stool samples [14]. Matsui and co-workers prepared cDNA from 10 μg of nucleic acid isolated from 1.5 g of infectious stool sample estimated to contain 10^5 to 10^6 virion particles per gram [9]. Vreede *et al.* cloned as little as 1 ng viral dsRNA using SISPA with 30 PCR cycles [24].

VARIATIONS AND ADAPTATIONS OF SISPA

VIDISCA

In 2004, a new human coronavirus, HCoV-NL63, was recovered using a variant of SISPA called

DOI: 10.1002/rmv

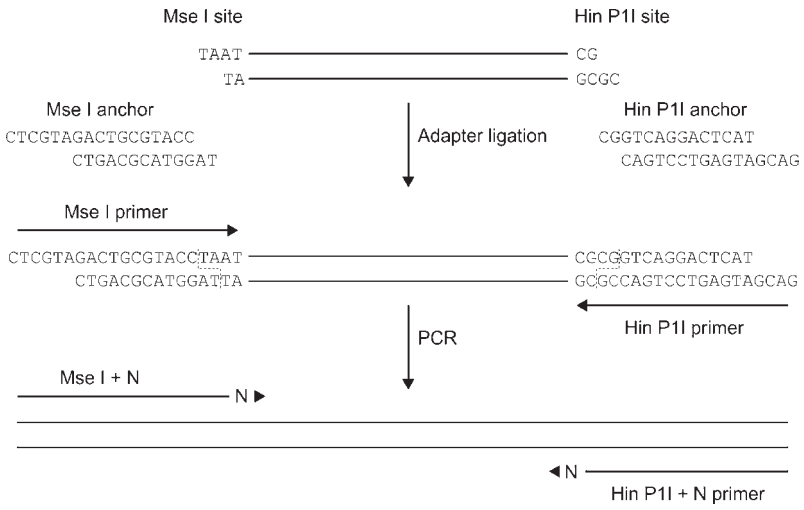

Figure 6. Anchors and primers used in the VIDISCA method. The starting template is converted to dsDNA as indicated in Figure 5 and is then digested with two restriction enzymes for example *Mse*I and *HinP1*I. The two anchors (adapter-linkers) are prepared by annealing together the top and bottom strand oligonucleotides, and they are then ligated onto the digested dsDNA fragments. The first PCR (e.g. 20 cycles) is done with the two primers, *Mse*I and *HinP1*I respectively, and the second PCR is performed with the same primer sequences except that they have an additional 3′-base (N). There are 16 of these primer combinations. The second PCR may be done with an input of 5 μL of the first reaction, using touchdown conditions (van der Hoek *et al*. [23])

*Vi*rus-*Dis*covery-*c*DNA-*A*FLP or VIDISCA [23] (Figure 6). It is based on the same principles as SISPA but uses two primers rather than one in the PCR amplification step, as is done in the amplified fragment length polymorphism (AFLP) technique [30], which is described below.

The DNA is digested with two frequently cutting restriction enzymes, for example *Mse*I and *HinPII*, both of which have four base pair recognition sites. This produces DNA molecules with *Mse*I and *HinP1*I overhangs at either end, as well as some with *Mse*I–*Mse*I and *HinP1*I–*HinP1*I overhangs. Only the *Mse*I and *HinP1*I fragments are amplified in the subsequent PCR as each adapter binds to one specific end of the DNA fragment, according to its complementary overhang. Two primers specific to each adapter are then used in an exponential amplification reaction by PCR. A second selective nested PCR amplification can be used to simplify the resultant PCR products from a DNA smear to specific bands. By extending the 3′ end of the primers by one to three nucleotides, a subset of the PCR products is generated [31]. There are 16 different possible primer combinations (4×4) if each primer is extended by only one nucleotide (i.e. *Mse*1 + N and *Hin*P1I + N). The use of two adapters and primers, and also the nested PCR step, makes VIDISCA more sensitive and specific than SISPA.

Other adaptations

Another SISPA variation was devised by Vreede *et al*. [24] as an adaptation of Lambden and Clarke's method [14] (shown in Figure 4). This variation allowed full-length copies of the dsRNA segments, 3–4 kb in size, of the African horse sickness virus (AHSV) genome to be cloned [24]. The basis of the method was the use of a primer with a 3′-poly(A) tail so that cDNA synthesis could be initiated with an oligo(dT) primer and the whole genome segment copied. The resultant cloned genome segments were sequenced and expressed *in vitro*. The method was developed further so that one-tube reactions could be performed for the adapter ligation, cDNA synthesis and PCR amplification [17]. A reduced number of PCR cycles were used (22–30 cycles compared to 30–35 of other methods), and a minimum of 1 ng dsRNA was needed to clone complete genomes.

Other modifications of SISPA include the incorporation of carrier mRNA, which may be especially helpful when trying to synthesise cDNA from pg amounts of mRNA. The addition of carrier mRNA (20 ng) was shown to increase the

DOI: 10.1002/rmv

number of recombinant colonies over background levels when as little as 20 pg of globin mRNA was used as starting RNA for cDNA synthesis [12].

An enrichment system has been devised using adapters with a biotin/streptavidin capture system for direct selection of the cDNA to be amplified [32]. Also, sequence-independent amplification of DNA has been used for the molecular cloning of specific microdissected DNA chromosomal regions [13]. Finally, a common epitope region of enteroviruses has been identified by SISPA followed by immunoscreening [33].

Amplified fragment length polymorphism

AFLP, a DNA fingerprinting technique, is essentially the same procedure as VIDISCA. Prior to amplification, the DNA is digested by restriction enzymes followed by ligation of double-stranded adapters to the termini of the DNA fragments [31,34]. Like VIDISCA, the selective amplification is achieved by the use of primers that extend into the restriction fragments, amplifying only those fragments in which the primer extensions match the nucleotides flanking the restriction sites. Using this method, subsets of restriction fragments may be visualised by PCR without any prior knowledge of nucleotide sequence. AFLP has been used to generate mRNA fingerprints in polyploid crop plants to identify deleted chromosomes [35]. It has also been used for the phylogenetic analysis of *Bacillus anthracis* [36] and *Escherichia coli* [37].

RANDOM PCR

Random PCR (rPCR), a technique similar to SISPA, uses a first primer with a 5′ end unique (e.g. 20) nucleotide universal sequence, containing restriction enzyme sites for subsequent cloning, followed by a degenerate hexa- or heptamer sequence at the 3′ end [38,39] (Figure 7). A subsequent PCR amplification step is carried out with a second, specific primer complementary to the 5′ universal region of the first random primer. This removes the need for an adapter ligation step, which can render SISPA inefficient. Random PCR can be used to detect both DNA and RNA viral genomes [40]. Random PCR was shown to amplify as little as 0.1 fg of MS2 phage RNA [38] or two copies of human chromosome 21 DNA [39]. Microdissected human chromosomal material was also amplified by random PCR, but with a primer with a random pentamer rather than hexa- or heptamer at its 3′ end [41]. HSV-1 was detected in a mouthwash sample from a patient with chronic fatigue syndrome using random PCR [40]. It has also been used to recover calicivirus RNA [42,43].

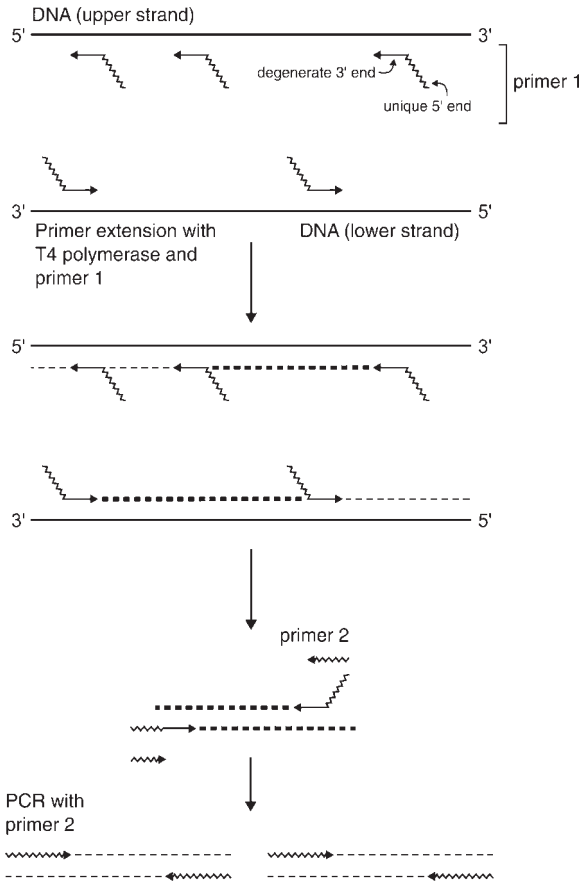

Figure 7. Random PCR. Random PCR can be used to amplify both DNA and RNA viral genomes. Here, a primer with a unique 5′ end sequence (indicated by the wavy line) and a fully degenerate 3′ end sequence is used in a PCR reaction to amplify viral DNA. Usually a lower annealing temperature or touchdown conditions are used. The degenerate part of the primer anneals to complementary DNA sequences which occur stochastically throughout the viral genome. The primer is extended using T4 polymerase. Then the generated double stranded sequences are separated by denaturation and complementary sequences hybridise to form finite portions of dsDNA with the first primer present at each end (as shown by the darker lines in the diagram). A primer representing only the unique sequence of the first primer is used for subsequent amplification of the fragments of the target (Stang *et al.* [40])

REPRESENTATIONAL DIFFERENCE ANALYSIS

The technique of RDA combines subtractive hybridisation with gene amplification to detect differ-

DOI: 10.1002/rmv

ences between two similar clinical samples [44]. In virology, the most likely samples to be compared are pre- and post-infection samples. The DNA isolated from the pre-infection sample acts as the 'driver' and is compared with the DNA isolated from the post-infection sample, the 'tester'. The two DNA samples are hybridised together so as to reduce common sequences, leaving mainly viral sequences for downstream analysis.

As Figure 8 shows, the genome complexity is reduced in both samples by a restriction enzyme digest, following which adapters are ligated only to the restriction fragments from the tester DNA. The two digested DNA samples are then combined, heated to melt the double-strands, and then cooled to allow the strands to anneal back together. Complementary sequences from the tester and driver DNA samples will hybridise together, while unique tester DNA sequences can only hybridise to each other. The ends of the DNA fragments are filled in, and then a primer complementary to the adapter sequence is used for PCR amplification. Heterogeneous annealed tester-driver fragments will undergo linear amplification because they have only one adapter sequence (from the tester DNA strand). The unique reannealed tester homogeneous fragments will undergo exponential amplification because they have two adapter sequences (from both tester DNA strands). The homogeneous reannealed driver DNA fragments have no adapter sequences, and so should not be amplified. Remaining single-stranded DNA fragments are digested with mung bean nuclease, which is specific for single-stranded DNA. By using the newly enriched tester amplicon as the starting material for each round, this subtraction/amplification procedure can be repeated several times. Sufficiently enriched in this way, the tester amplicons can then either be cloned for library construction, used as a probe source for library screening, or sequenced directly. Two highly matched nucleic acid sources are necessary for successful recovery of difference products, which severely limits the use of RDA. Also, amplification of sequences from both DNAs needs to occur equally well, but this cannot be guaranteed if the DNAs are amplified in separate tubes. However, RDA has been used to identify a gamma herpes virus [45], hepatitis G virus [46] and an anellovirus, Torque Teno virus (TTV) [47]. A balanced PCR method involving PCR amplification of both DNA samples in one tube has been described [48]. This has been claimed to eliminate differences in PCR amplification efficiency between the tester and driver samples.

Modifications of RDA

Hu *et al*. [49] described a modified RDA method for the identification of unknown viral agents. Large quantities of cDNA were made by a universal long-PCR method. Nested-PCR-based subtractive hybridisation, and the removal of single-stranded DNA, was then used to give a population of DNA that could be cloned to give a cDNA library. Hepatitis C virus was used as a control tester sample to validate the system, and it was successfully found in the cDNA library. As of June 2006, however, we could not find any reference to the use of this method for the discovery of any new virus.

Another method that has been described for pathogen discovery involves high-throughput sequencing of cDNA libraries made from infected tissues [50]. A recombinant cDNA library from RNA was prepared from a post-transplant lymphoproliferative disorder tissue sample. The sequences obtained from the clones were compared with human genome sequences by computational BLAST searching, and those that did not match the human genome were further investigated. After sequencing, a cDNA library of 27 840 transcripts, 32 transcripts were identified as non-human: of these, 10 were proven to be from EBV. Although we are not aware that this method has been used to discover any novel viruses, it is noteworthy that high-throughput or shotgun sequencing has been used to identify novel microorganism sequences from both seawater and biofilms [51,52]. The methods that were used in these two shotgun-sequencing studies could be applied in other contexts, including those associated with viral disease.

DIFFERENTIAL DISPLAY

Differential display (DD) is a powerful method used to investigate gene expression, specifically to look for a difference in the expression of mRNAs between two closely related samples [53,54]. In the context of infectious diseases these might be pre- and post-viral infection specimens that is RNA from a healthy tissue and RNA from a diseased tissue. RNA extracted from the sources

DOI: 10.1002/rmv

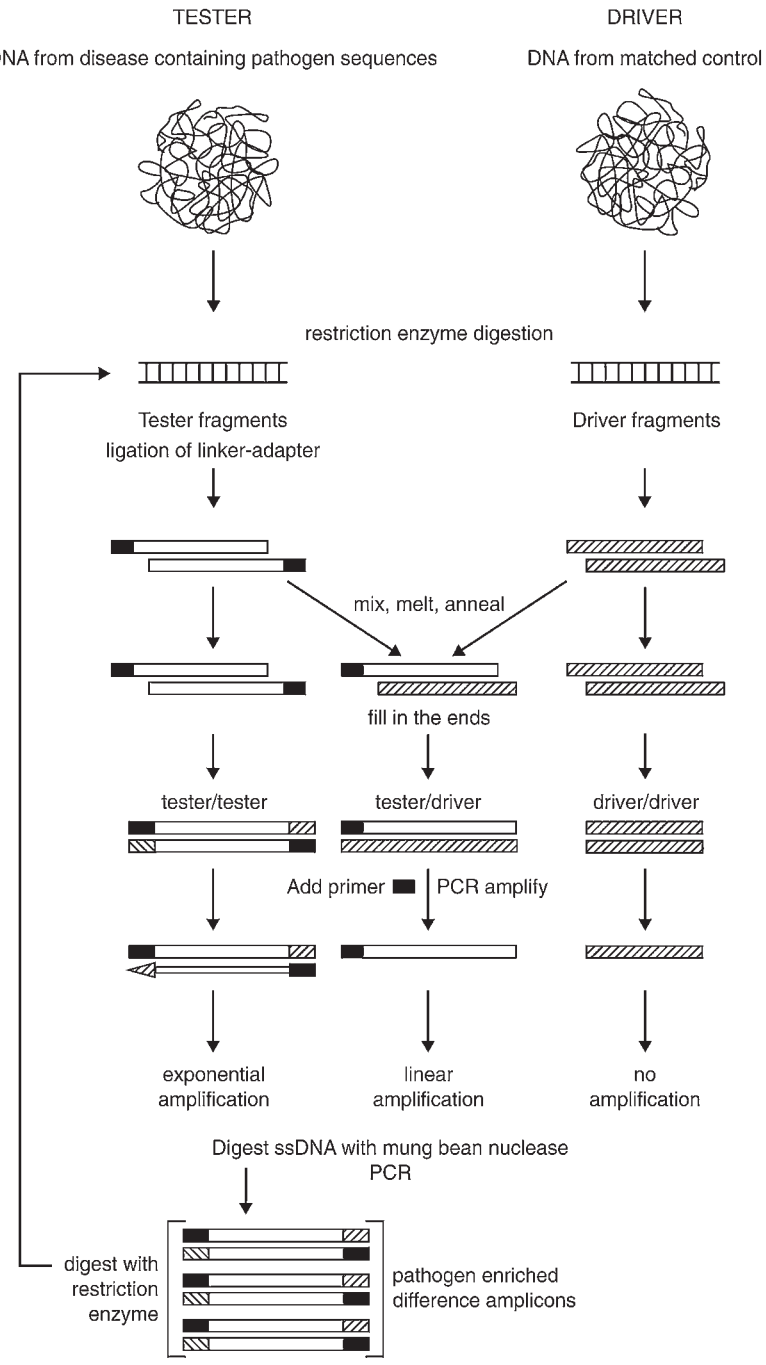

Figure 8. Representational Difference Analysis. For RDA, two DNA sources are needed, the tester and the driver. They differ only in that the tester contains pathogen sequences, while the driver does not. The driver is used at a higher concentration that the tester, to *drive* the reaction. The DNA samples are digested with a restriction enzyme (e.g. *Dpn*I). A linker/adapter (to provide the primer sequence in further PCR) is added *only* to the tester DNA digest. The two DNA populations are mixed, heated and annealed to form three kinds of molecules: tester/tester sequences; hybrids of tester and driver; driver/driver sequences. As there is an excess of driver DNA, the tester/tester molecules should be enriched for pathogen sequences because the tester non-pathogen sequences will hybridise to the corresponding driver sequences. The ends of the re-annealed DNA are filled in and then amplified by PCR with a primer specific for the linker/adapter sequence. The tester/tester molecules with the pathogen sequence should be preferentially and exponentially amplified. Nuclease digestion is used to remove unwanted ssDNA and further PCR is performed. More rounds of this procedure may be carried out by combining the resultant pathogen-enriched difference amplicons with an excess of driver DNA restriction enzyme fragments (Lisitsyn *et al.* [44])

DOI: 10.1002/rmv

to be compared is separately reverse transcribed with an 11-mer oligo dT-primer with a C, A or G base at its 3′ end, and AAGC at its 5′ end. The cDNA from this reaction is used as a template in a PCR with the same oligo dT-primer and with a 13-mer primer with a random sequence. The reaction is carried out in the presence of radiolabelled dATP that becomes incorporated into the PCR amplicons. The PCR amplicons are separated by electrophoresis on a denaturing polyacrylamide gel and visualised by autoradiography. It is assumed that any additional bands in the amplified RNA from the infected tissue are derived from viral mRNA. A picornavirus infecting honeybees has been identified in this way by DD, by comparing mRNA expression in brains of aggressive workers with that in the brains of nurse bees and foragers [55]. The workers were found to be infected with the picornavirus and it was suggested that this caused their aggressive behaviour.

WHOLE GENOME AMPLIFICATION

Sequence-independent amplification should be a means not only for identification of viral nucleic acid of unknown sequence, but also for that at low concentration. Some clinical samples may harbour only very few molecules of the unknown genomic viral nucleic acid, and sequence-independent amplification may not be sufficiently sensitive to reveal it. This problem arises because, in general, PCR amplification methods using generic primers are much less sensitive than methods using sequence-specific primers.

To overcome this problem, methods for amplification of all the nucleic acid can be applied after the sample has been sufficiently enriched for viral nucleic acid, as previously described, but prior to carrying out the sequence-independent amplification methods. Such techniques are called whole genome amplification (WGA) methods and include multiple displacement amplification (MDA), omniplex amplification, degenerate oligonucleotide primed PCR (DOP-PCR) and primer extension preamplification PCR (PEP-PCR).

Multiple displacement amplification

Of the several methods available for amplifying whole genomes, the most useful one may be MDA, which is available as a commercial kit (GE Healthcare UK Ltd., Little Chalfont, UK). DNA from clinical samples can be directly amplified using MDA without the need for prior purification.

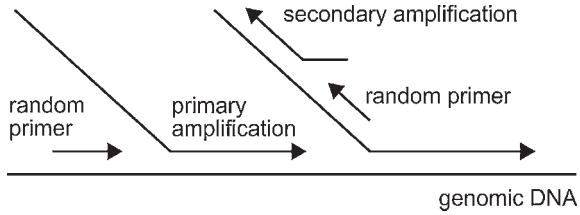

Figure 9. Multiple displacement amplification.MDA is accomplished with random primers and Φ29 DNA polymerase for 16–18 hours at 30°C. Primary amplification occurs when the polymerase replicates the genomic DNA using the random primers, and secondary amplification occurs on the displaced product strands (Dean *et al.* [56])

MDA exploits the high processivity of bacteriophage Φ29 polymerase, generating products of over 10 kb in size (Figure 9) [56]. Amplification occurs by strand-displacement with random exonuclease resistant hexamers during a 16 to 18 hour isothermal incubation at 30°C. The polymerase has an error rate of 1 in 10^6 to 10^7 nucleotides, which compares well with *Taq* polymerase, which has an error rate of 3 in 10^4 nucleotides. One significant property of MDA is that both large fragments and amounts of DNA are generated in one step, and this DNA can be used in further rounds of reamplification. The yield of DNA is consistent and is representative of the whole genome, as was demonstrated by single nucleotide polymorphism assays and comparative genome hybridisation [56].

There are contradictory reports on the amount of starting DNA needed for a successful MDA reaction. Barker *et al*. [57] claim that consistent DNA yields are observed regardless of the amount of starting material, and that 75- to 80-fold amplification is achieved. Dean *et al*. [56] suggest that 1 to 10 copies of human genomic DNA can generate 20 to 30 μg of product. Like Barker, they claim that clinical samples that differ in quality and concentration produce similar and reproducible yields after MDA. By contrast, Lovmar *et al*. [58] report that successful amplification does depend on the amount of starting material, and they recommend using 3 ng or 1000 genome equivalents.

Omniplex amplification

Omniplex is another WGA method, and also is available as a commercial kit (Sigma–Aldrich Com-

DOI: 10.1002/rmv

pany Ltd., Poole, UK). Genomic DNA is randomly fragmented generating a restricted size range of products (from 100 to 1000 bp), which are converted into a library of inherently amplifiable DNA molecules. A high-fidelity polymerase is used to generate about 1000-fold amplification [57]. The amplification can be monitored in real-time and, like MDA, generates very reproducible amplification across the genome. The main advantage of the Omniplex method is that it can be used for both degraded and intact DNA, and the resulting DNA products can be archived and repeatedly reamplified.

Other whole genome amplification methods

There are two other methods for whole genome amplification: DOP-PCR and PEP-PCR [59]. These were initially designed for amplification from single haploid cells for pre-implantation genetic disease diagnosis. There are caveats to both methods, however, as they generate non-specific amplification artefacts and short unusable products, and there is incomplete coverage of the whole genome. Both methods use degenerate 15-mer PCR primers, which could lead to artificial sequences occurring in the amplification product.

RNA AMPLIFICATION

Whole mRNA amplification is also possible and there are several commercial kits available. These involve a series of enzymatic reactions resulting in linear amplification. Most of these are based on the antisense RNA (aRNA) amplification method first described by Van Gelder and Eberwine [60]. This was developed for amplifying and labelling exceedingly small amounts of mRNA for hybridisation to microarrays, usually for gene expression studies. Unlike exponential RNA amplification methods, such as RT-PCR, aRNA amplification maintains representation of the starting mRNA population [61]. The procedure begins with total or just poly(A) RNA that is reverse transcribed using a primer containing both oligo(dT) and a T7 RNA polymerase promoter sequence (Figure 10). After first-strand synthesis, the reaction is treated with RNase H to cleave the mRNA into small fragments. These small RNA

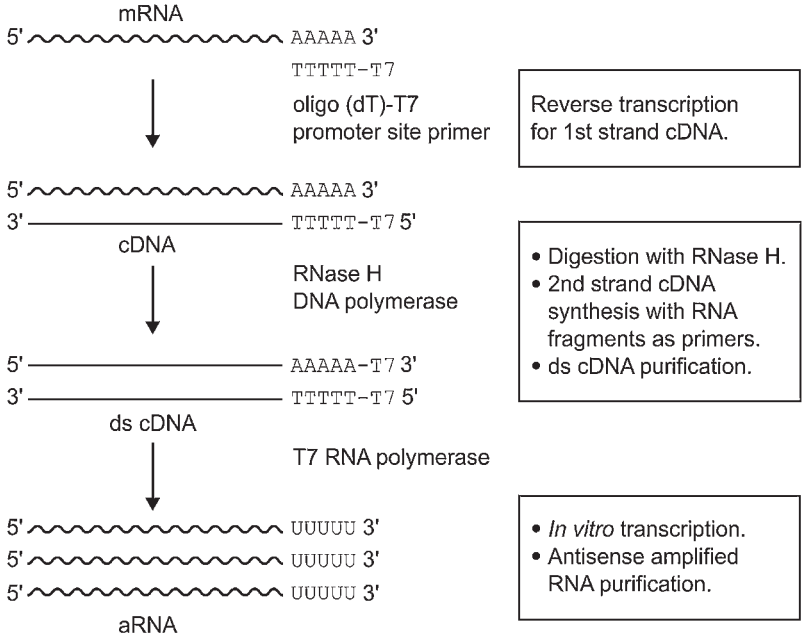

Figure 10. RNA amplification procedure. RNA amplification is achieved by making a dsDNA template of the target, and then transcribing ssRNA copies from it. An oligo(dT)-T7 promoter site primer is used to initiate cDNA synthesis from mRNA by the enzyme reverse transcriptase. The resulting hybrid of cDNA and RNA is digested with RNase H and the RNA fragments serve as primers for the second strand cDNA synthesis using DNA polymerase. The double-stranded cDNA is then purified and acts as the template for the *in vitro* transcription using T7 RNA polymerase. During the generation of the cDNA, the T7 promoter site has been incorporated into the dsDNA and serves as the initiation site for T7 RNA polymerase. The resulting antisense amplified RNA is then purified for downstream applications

DOI: 10.1002/rmv

fragments serve as primers during a second-strand synthesis reaction that produces a double-stranded cDNA template for transcription. Contaminating rRNA, mRNA fragments and primers are removed and the cDNA template is then used in an *in vitro* transcription reaction to produce linearly amplified aRNA.

VIRUS CULTURE

One advantage of sequence-independent amplification is that it is culture independent. This does not, however, eliminate the possibility of first inoculating the specimen into cell culture to see if any cpe can be observed. If it were, it may be assumed to be due to the growth of an unknown virus, and would hence provide a means of amplifying the genome. Even in the absence of cpe, some viral growth may have occurred.

By way of example, in order to characterise HCoV-NL63, the nasopharyngeal aspirate specimen was inoculated onto *Cynomolgus* monkey kidney cells, human fetal lung fibroblasts and HeLa cells [23]. A cytopathic effect was detected exclusively with the monkey kidney cells, with a more pronounced cpe being observed on passage onto another monkey kidney cell line, LLC-MK2. VIDISCA was then applied to the supernatant of this culture and subsequently HCoV-NL63 was identified. A specific PCR was designed to confirm the presence of HCoV-NL63, both in the cell culture and the clinical sample, and was then used as a diagnostic PCR in further clinical samples.

In one series of experiments [40], supernatants from positive virus cultures were used to develop a random PCR protocol. This protocol was validated using a mouth wash sample from a patient suffering from chronic fatigue syndrome. The clinical sample created a cpe in HeLa cells and, by cloning and sequencing, HSV-1 was shown to be present.

Novel cell lines are being developed that may be useful for growth of unknown viruses. These include semi-differentiated stem cells and cells deficient in interferon signalling that lack the transcription factor, the signal transducer and activator of transcription 1 (STAT1) [62]. Transgenic mice may also be used to support the growth of novel pathogens [63].

CHARACTERISING THE PRODUCTS OF SEQUENCE-INDEPENDENT AMPLIFICATION

Sequence analysis

The majority of viruses discovered by sequence-independent amplification have been characterised by cloning and sequencing the PCR amplicons, and then using BLAST programs and phylogenetic analysis to analyse the sequence data [20,22,23,40]. Key to this approach, however, is that the unknown viruses have some homology to known viruses present in the databases. It has, nevertheless, proven possible to identify viruses that are only very distantly related to known viruses for example the novel parvoviruses, bocavirus and Parv4 [22,25].

Specific programmes have been developed for the analysis of sequence data from defined sources. For example, for bacteriophage sequence data, there is the PHAge Communities from Contig Spectrum (PHACCS) programme [64]. This web-based tool was designed to predict viral community structure and diversity using sequence contigs generated from viral shotgun libraries. The sequence data can be mathematically analysed to increase understanding of viral ecology and population dynamics. For example, PHACCS has been used to verify that phage biodiversity is greater than in any previously observed community. To our knowledge, no programmes have been designed to calculate this for human viruses.

Microarrays

Wang *et al*. [26] designed comprehensive DNA microarrays for viral discovery and sequence recovery. Their second-generation DNA microarray consists of 70-mer oligonucleotides derived from every fully sequenced reference viral genome in Genbank (as of 15 August 2002) [27]. To maximise the probability of characterising unknown and unsequenced members of existing families, the most highly conserved 70-mers from each virus were selected. A mean number of ten 70-mers was chosen for each virus, totalling approximately 10 000 oligonucleotides from about 1000 viruses.

This pan-viral array was used as part of the global effort to identify the novel coronavirus (SARS-CoV) associated with severe acute respiratory syndrome (SARS). Prior to identification, the virus

DOI: 10.1002/rmv

was cultured in Vero cells from a patient suffering from SARS [65]. Total nucleic acid was purified from the culture and then amplified and hybridised to the array. For further characterisation of the virus, approximately 1 kb of the viral genome was cloned and sequenced by physically recovering viral sequences hybridised to the individual array elements.

In addition, a random PCR amplification strategy was developed in order to amplify viral nucleic acid to be identified using the microarray [26]. This was tested using human respiratory specimens from which multiple viruses were detected.

Other viral specific microarrays have been reported against which PCR amplicons from sequence-independent amplification reactions could be matched and identified. Boriskin *et al*. [66] developed a high-resolution, low-density diagnostic DNA microarray specific for central nervous system viral infections. It has been utilised for the detection of multiplex PCR-amplified viruses in CSF and non-CSF specimens. The array contains 38 gene targets for 13 viral causes of meningitis and encephalitis, and was tested using 41 clinical specimens. The clinical sensitivity, specificity and negative and positive predictive values were determined to be 93%, 100% 100%, and 83% respectively when the results from these tests were compared with those of single virus PCR. However, Boriskin *et al*. reported that the interpretation of a negative result is difficult because it is affected by assay sensitivity, low viral genome number and sample-specific inhibition. Other arrays that have been described include an oligonucleotide microarray for the rapid detection and serotyping of acute respiratory disease-associated adenoviruses [67], and a DNA probe array for the simultaneous identification of herpesviruses, enteroviruses and flaviviruses [68].

An array-based pathogen chip has been developed for the detection of viral RNA or DNA relevant to the pathologies of the central nervous system [69]. Open reading frames with highly conserved and heterogenic sequence regions within viral families were used to design a total of 715 unique oligonucleotides (60-mers), which represent approximately 100 pathogens. Viral genes representing different stages of pathogen infection were also put on the chip to potentially characterise the stage of the viral infection. The array was tested with six post-mortem brain samples from which human CMV and HSV were detected.

Microarrays that consist of a comprehensive and all encompassing set of sequence probes have much promise for virus characterisation, provided the viral DNA or RNA present in the specimen can be amplified and labelled by multiplex, generic or consensus PCR [70], or by the methods described in this review, and provided it has sufficient homology to the known viral sequences.

Other methods for post-amplification analysis

Reyes *et al*. [8] used Southern blot hybridisation to quantify the amount of virus production using SISPA. Bovine leukaemia virus cDNA pre- and post- SISPA was analysed by agarose gel electrophoresis, Southern blotted and probed with an *env* gene fragment specific for bovine leukaemia virus. The probe only bound to the cDNA subjected to SISPA.

To detect the common epitope region of enteroviruses, cDNA libraries constructed from SISPA methods were immunoscreened using anti-enterovirus guinea pig antisera and the antisera from patients with aseptic meningitis [33]. Through repeated immunoscreening, 82 immunopositive cDNA clones were selected and sequenced. Of these clones, 31 were located on the upstream region of VP1.

KOCH'S PROBLEM

Koch's postulates are well known as a set of criteria that have to be fulfilled by a microbe for it to be proven as the causative agent of disease. They were modified in 1937 for viruses by Rivers and, lately, SARS-CoV has been shown to fulfil them, but at the expense of infecting macaque monkeys [71,72]. Arguably, the Koch-Rivers' postulates need to be adapted for pathogen identification by sequence discovery. Until that happens, however, care will need to be taken in interpreting the presence of viral sequences in a clinical specimen as proof of causation of disease. The most obvious strategy following sequence discovery is to make a specific PCR for the new virus and use it to test many specimens from different individuals with that disease and as a control, individuals without the disease. If there are many positive results, there is a high likelihood that the virus is pathogenic for that disease. It needs to be

DOI: 10.1002/rmv

remembered that finding the virus is only the first step in establishing a cause for a disease.

CONCLUDING REMARKS

At present, SISPA and other methods for amplification of viral genomes of unknown sequence are complicated and prone to amplification of sequence artefacts with spurious results. These methods have, however, proved their worth in the recovery of previously unknown viruses (see Table 1) and, if more reliable, robust and reproducible versions of them can be developed, they are likely to find widespread application. They have great potential when combined with suitable end-stage detection methods such as microarrays and high throughput sequencing for the identification of candidate pathogen sequences in clinical specimens.

When investigating diseases of unknown aetiology, standard virological techniques should not be neglected in favour of molecular biological methods simply because the latter are thought to be more fashionable. Cell culture has been mentioned above, but electron microscopy, serology, immunofluorescence and standard PCR tests should also be used where applicable. Perhaps the starting point for any molecular investigation of a disease of unknown aetiology should be the use of PCR with consensus primers for the most likely virological suspects.

Viruses may often be identified through their proteins, specifically through antigen–antibody interactions. In the context of virus discovery, antibodies shown to be absent pre-disease but present in the sera from patients recovering from disease may be diagnostic. Convalescent sera may also be used to purify viruses from complex mixtures, prior to the extraction and amplification of their genome. More ambitiously, it may be possible to reconstruct recombinant antibodies from immune cells collected from patients with disease, and then use them to identify antigens from the virus causing that disease. This approach has been demonstrated to be feasible for subacute sclerosing panencephalitis and measles virus, and is being developed for multiple sclerosis [73,74].

In conclusion, any virus discovery project should exploit conventional virological methods as well as molecular techniques for the amplification of unknown sequences. Novel cell culture and antibody-based methods should not, therefore, be neglected. Epidemiological evidence should support any proposed link between the disease being studied and an infectious cause. Lastly, attention should be given to the collection and selection of satisfactory samples. In particular, the timing of the samples is critical in virus discovery.

ACKNOWLEDGEMENTS

We thank Dr David Brown, Dr Philip Mortimer and Dr Kevin Brown for helpful suggestions, corrections and comments.

REFERENCES

1. Cotmore SF, Tattersall P. Characterisation and molecular cloning of a human parvovirus genome. *Science* 1984; **226**: 1161–1165.
2. Clewley JP. Detection of human parvovirus using a molecularly cloned probe. *J Med Virol* 1985; **15**: 173–181.
3. Choo Q-L, Kuo G, Weiner A, Overby LR, Bradley DW, Houghton M. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science* 1989; **244**: 359–362.
4. Clewley JP, Lewis JCM, Brown DWG, Gadsby EL. A novel simian immunodeficiency virus (SIVdrl) pol sequence from the drill monkey, Mandrillus leucophaeus. *J Virol* 1998; **72**: 10305–10309.
5. Donehower LA, Bohannon RC, Ford RJ, Gibbs RA. The use of primers from highly conserved pol regions to identify uncharacterized retroviruses by the polymerase chain reaction. *J Virol Meth* 1990; **28**: 33–46.
6. Wichman HA, Van Den Bussche RA. In search of retrotransposons: exploring the potential of the PCR. *BioTechniques* 1992; **13**: 258–264.
7. Muerhoff A, Leary T, Desai S, Mushahwar I. Amplification and subtraction methods and their application to the discovery of novel human viruses. *J Med Virol* 1997; **53**: 96–103.
8. Reyes GR, Kim JP. Sequence-independent, single-primer amplification (SISPA) of complex DNA populations. *Mol Cell Probes* 1991; **5**: 473–481.
9. Matsui SM, Kim JP, Greenberg HB, *et al*. The isolation and characterization of a Norwalk virus-specific cDNA. *J Clin Invest* 1991; **87**: 1456–1461.
10. Matsui SM, Kim JP, Greenberg HB, *et al*. Cloning and characterization of human astrovirus immunoreactive epitopes. *J Virol* 1993; **67**: 1712–1715.
11. Linnen J, Wages J, Jr., Zhang-Keck Z-Y, *et al*. Molecular cloning and disease association of hepatitis G virus: a transfusion-transmissible agent. *Science* 1996; **271**: 505–508.
12. Akowitz A, Manuelidis L. A novel cDNA/PCR strategy for efficient cloning of small amounts of undefined RNA. *Gene* 1989; **81**: 295–306.

DOI: 10.1002/rmv

13. Johnson DH. Molecular cloning of DNA from specific chromosomal regions by microdissection and sequence-independent amplification of DNA. *Genomics* 1990; **6**: 243–251.
14. Lambden PR, Cooke SJ, Caul EO, Clarke IN. Cloning of noncultivatable human rotavirus by single primer amplification. *J Virol* 1992; **66**: 1817–1822.
15. Lambden PR, Clarke IN. Cloning of Viral Double-Stranded RNA genomes by single primer amplification. In *Cloning of Viral Double-Stranded RNA Genomes by Single Primer Amplification*. Academic Press: Minneapolis, 1995; 359–372.
16. Chen Z, Lambden PR, Lau J, Caul EO, Clarke IN. Human group C rotavirus: completion of the genome sequence and gene coding assignments of a non-cultivatable rotavirus. *Virus Res* 2002; **83**: 179–187.
17. Potgieter AC, Steele AD, van Dijk AA. Cloning of complete genome sets of six dsRNA viruses using an improved cloning method for large dsRNA genes. *J Gen Virol* 2002; **83**: 2215–2223.
18. Tsunemitsu H, Kamiyama M, Kawashima K, *et al*. Molecular characterization of the major capsid protein VP6 of bovine group B rotavirus and its use in seroepidemiology. *J Gen Virol* 2005; **86**: 2569–2575.
19. Allander T, Emerson SU, Engle RE, Purcell RH, Bukh J. A virus discovery method incorporating DNase treatment and its application to the identification of two bovine parvovirus species. *Proc Natl Acad Sci USA* 2001; **98**: 11609–11614.
20. Allander T, Emerson SU, Purcell RH, Bukh J. Cloning of unknown virus sequences by DNase treatment and sequence independent single primer amplification. In *Cloning of Unknown Virus Sequences by DNase Treatment and Sequence Independent Single Primer Amplification*. CRC Press: Boca Raton, 2004; 337–345.
21. Breitbart M, Rohwer F. Method for discovering novel DNA viruses in blood using viral particle selection and shotgun sequencing. *Biotechniques* 2005; **39**: 729–736.
22. Allander T, Tammi MT, Eriksson M, Bjerkner A, Lindell AT, Andersson B. Cloning of a human parvovirus by molecular screening of respiratory tract samples. *Proc Natl Acad Sci USA* 2005; **102**: 12891–12896.
23. van der Hoek L, Pyrc K, Jebbink MF, *et al*. Identification of a new human coronavirus. *Nature Med* 2004; **10**: 368–373.
24. Vreede FT, Cloete M, Napier GB, van Dijk AA, Viljoen GJ. Sequence-independent amplification and cloning of large dsRNA virus genome segments by poly(dA)-oligonucleotide ligation. *J Virol Meth* 1998; **72**: 243–247.
25. Jones MS, Kapoor A, Lukashov VV, Simmonds P, Hecht F, Delwart E. New DNA viruses identified in patients with acute viral infection syndrome. *J Virol* 2005; **79**: 8230–8236.
26. Wang D, Coscoy L, Zylberberg M, *et al*. Microarray-based detection and genotyping of viral pathogens. *Proc Natl Acad Sci USA* 2002; **99**: 15687–15692.
27. Wang D, Urisman A, Liu Y-T, *et al*. Viral discovery and sequence recovery using DNA microarrays. *PLoS Biology* 2003; **1**: e2.
28. Vora GJ, Meador CE, Stenger DA, Andreadis JD. Nucleic acid amplification strategies for DNA microarray-based pathogen detection. *Appl Environ Microbiol* 2004; **70**: 3047–3054.
29. Reyes GR, Bradley DW, Lovett M. New strategies for isolation of low abundance viral and host cDNAs: application to cloning of the hepatitis E virus and analysis of tissue-specific transcription. *Sem Liver Dis* 1992; **12**: 289–300.
30. Bachem CWB, van der Hoeven RS, de Bruijn SM, Vreugdenhil D, Zabeau M, Visser RGF. Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development. *Plant J* 1996; **9**: 745–753.
31. Vos P, Hogers R, Bleeker M, *et al*. AFLP: a new technique for DNA fingerprinting. *Nuc Acids Res* 1995; **23**: 4407–4414.
32. Morgan JG, Dolganov GM, Robbins SE, Hinton LM, Lovett M. The selective isolation of novel cDNAs encoded by the regions surrounding the human interleukin 4 and 5 genes. *Nuc Acids Res* 1992; **20**: 5173–5179.
33. Shin S-Y, Kim K-S, Lee Y-S, *et al*. Identification of enteroviruses by using monoclonal antibodies against a putative common epitope. *J Clin Microbiol* 2003; **41**: 3028–3034.
34. Arnold C. Fluorescent amplified fragment length polymorphism (FAFLP) genotyping. In *Fluorescent Amplified Fragment Length Polymorphism (FAFLP) genotyping*. CRC Press: Boca Raton, 2004; 155–161.
35. Money T, Reader S, Qu LJ, Dunford RP, Moore G. AFLP-based mRNA fingerprinting. *Nuc Acids Res* 1996; **24**: 2616–2617.
36. Keim P, Kalif A, Schupp J, *et al*. Molecular evolution and diversity in Bacillus anthracis as detected by amplified length polymorphism markers. *J Bacteriol* 1997; **179**: 818–824.
37. Arnold C, Metherell L, Willshaw G, Maggs A, Stanley J. Predictive fluorescent amplified-fragment length polymorphism analysis of Escherichia coli: high-resolution typing method with phylogenetic significance. *J Clin Microbiol* 1999; **37**: 1274–1279.

DOI: 10.1002/rmv

38. Froussard P. A random-PCR method (rPCR) to construct whole cDNA library from low amounts of RNA. *Nuc Acids Res* 1992; **20**: 2900.
39. Grothues D, Cantor CR, Smith CL. PCR amplification of megabase DNA with tagged random primers (T-PCR). *Nuc Acids Res* 1993; **21**: 1321–1322.
40. Stang A, Korn K, Wildner O, Uberla K. Characterization of virus isolates by particle-associated nucleic acid PCR. *J Clin Microbiol* 2005; **43**: 716–720.
41. Bohlander SK, Espinosa R, 3rd, Le Beau MM, Rowley JD, Diaz MO. A method for the rapid sequence-independent amplification of microdissected chromosomal material. *Genomics* 1992; **13**: 1322–1324.
42. Dingle KE, Lambden PR, Caul EO, Clarke IN. Human enteric Caliciviridae: the complete genome sequence and expression of virus-like particles from a genetic group II small round structured virus. *J Gen Virol* 1995; **76**: 2349–2355.
43. Liu BL, Lambden PR, Gunther H, Otto P, Elschner M, Clarke IN. Molecular characterization of a bovine enteric calicivirus: relationship to the Norwalk-like viruses. *J Virol* 1999; **73**: 819–825.
44. Lisitsyn N, Lisitsyn N, Wigler M. Cloning the differences between two complex genomes. *Science* 1993; **259**: 946–951.
45. Chang Y, Cesarman E, Pessin MS, *et al*. Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. *Science* 1994; **266**: 1865–1869.
46. Simons JN, Pilot-Matias TJ, Leary TP, *et al*. Identification of two flavivirus-like genomes in the GB hepatitis agent. *Proc Natl Acad Sci USA* 1995; **92**: 3401–3405.
47. Nishizawa T, Okamoto H, Konishi K, Yoshizawa H, Miyakawa Y, Mayumi M. A novel DNA virus (TTV) associated with elevated transaminase levels in post-transfusion hepatitis of unknown etiology. *Biochem Biophys Res Commun* 1997; **241**: 92–97.
48. Makrigiorgos GM, Chakrabarti S, Zhang Y, Kaur M, Price BD. A PCR-based amplification method retaining the quantitative difference between two complex genomes. *Nature Biotechnol* 2002; **20**: 936–939.
49. Hu Y, Hirshfield I. Rapid approach to identify an unrecognized viral agent. *J Virol Meth* 2005; **127**: 80–86.
50. Xu Y, Stange-Thomann N, Weber G, *et al*. Pathogen discovery from human tissue by sequence-based computational subtraction. *Genomics* 2003; **81**: 329–335.
51. Venter JC, Remington K, Heidelberg JF, *et al*. Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 2004; **304**: 66–74.
52. Tyson GW, Chapman J, Hugenholtz P, *et al*. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* 2004; **428**: 37–43.
53. Liang P, Pardee AB (eds). *Differential Display Methods and Protocols*. Humana Press; Totowa, New Jersey, 1997.
54. Bosch I, Warke RV, Fournier MF, Magee D, Melichar H. Applications of differential display in infectious diseases. In *Applications of Differential Display in Infectious Diseases*. CRC Press: Boca Raton, 2004; 277–285.
55. Fujiyuki T, Takeuchi H, Ono M, *et al*. Novel insect picorna-like virus identified in the brains of aggressive worker honeybees. *J Virol* 2004; **78**: 1093–1100.
56. Dean FB, Hosono S, Fang L, *et al*. Comprehensive human genome amplification using multiple displacement amplification. *Proc Natl Acad Sci USA* 2002; **99**: 5261–5266.
57. Barker DL, Hansen MST, Faruqi AF, *et al*. Two methods of whole-genome amplification enable accurate genotyping across a 2320-SNP linkage panel. *Genome Res* 2004; **14**: 901–907.
58. Lovmar L, Fredriksson M, Liljedahl U, Sigurdsson S, Syvanen A-C. Quantitative evaluation by minisequencing and microarrays reveals accurate multiplexed SNP genotyping of whole genome amplified DNA. *Nucl Acids Res* 2003; **31**: e129.
59. Zhong XY, Li Y, Holzgreve W, Hahn S. Primer extension preamplification (PEP) of single cells: efficiency and bias. In *Primer Extension Preamplification (PEP) of Single Cells: Efficiency and Bias*. CRC Press: Boca Raton, 2004; 361–368.
60. Van Gelder RN, von Zastrow ME, Yool A, Dement WC, Barchas JD, Eberwine JH. Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc Natl Acad Sci USA* 1990; **87**: 1663–1667.
61. Poirier GM, Erlander MG. Postdifferential display: parallel processing of candidates using small amounts of RNA. *Methods* 1998; **16**: 444–452.
62. Barreca C, O'Hare P. Suppression of herpes simplex virus 1 in MDBK cells via the interferon pathway. *J Virol* 2004; **78**: 8641–8653.
63. Lassnig C, Kolb A, Strobl B, Enjuanes L, Muller M. Studying human pathogens in animal models: Fine tuning the humanized mouse. *Transgenic Res* 2005; **14**: 803–806.
64. Angly F, Rodriguez-Brito B, Bangor D, *et al*. PHACCS, an online tool for estimating the structure and diversity of uncultured viral communities using metagenomic information. *BMC Bioinformatics* 2005; **6**: 41.
65. Ksiazek TG, Erdman D, Goldsmith CS, *et al*. A novel coronavirus associated with severe acute respiratory syndrome. *N Engl J Med* 2003; **348**: 1953–1966.

DOI: 10.1002/rmv

66. Boriskin YS, Rice PS, Stabler RA, *et al*. DNA microarrays for virus detection in cases of central nervous system infection. *J Clin Microbiol* 2004; **42**: 5811–5818.
67. Lin B, Vora GJ, Thach D, *et al*. Use of oligonucleotide microarrays for rapid detection and serotyping of acute respiratory disease-associated adenoviruses. *J Clin Microbiol* 2004; **42**: 3232–3239.
68. Korimbocus J, Scaramozzino N, Lacroix B, Crance JM, Garin D, Vernet G. DNA probe array for the simultaneous identification of herpesviruses, enteroviruses, and flaviviruses. *J Clin Microbiol* 2005; **43**: 3779–3787.
69. Conejero-Goldberg C, Wang E, Yi C, *et al*. Infectious pathogen detection arrays: viral detection in cell lines and postmortem brain tissue. *Biotechniques* 2005; **39**: 741–751.
70. Clewley JP. A role for arrays in clinical virology: fact or fiction? *J Clin Virol* 2004; **29**: 2–12.
71. Fouchier RAM, Kuiken T, Schutten M, *et al*. Aetiology: Koch's postulates fulfilled for SARS virus. *Nature* 2003; **423**: 240–240.
72. Osterhaus AD, Fouchier RA, Kuiken T. The aetiology of SARS: Koch's postulates fulfilled. *Philos Trans R Soc Lond B Biol Sci* 2004; **359**: 1081–1082.
73. Burgoon MP, Keays KM, Owens GP, *et al*. Laser-capture microdissection of plasma cells from subacute sclerosing panencephalitis brain reveals intrathecal disease-relevant antibodies. *Proc Natl Acad Sci USA* 2005; **102**: 7245–7250.
74. Gilden DH. Viruses and multiple sclerosis. *J Am Med Assoc* 2001; **286**: 3127–3129.
75. Reyes GR, Purdy MA, Kim JP, *et al*. Isolation of a cDNA from the virus responsible for enterically transmitted non-A, non-B hepatitis. *Science* 1990; **247**: 1335–1339.
76. Tam AW, Smith MM, Guerra ME, *et al*. Hepatitis E virus (HEV): molecular cloning and sequencing of the full- length viral genome. *Virology* 1991; **185**: 120–131.
77. Chang Y, Cesarman E, Pessin MS, *et al*. Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. *Science* 1994; **266**: 1865–1869.

DOI: 10.1002/rmv