amongst others, that are expected to affect intra and inter assays. These variations may be minimised by increased staff training and the use of high-quality automated precision equipment.

The modern microbiologist has an impressive toolkit with molecular and non-molecular tests to choose from. Sometimes the choice of test for a given pathogen varies between different clinical scenarios, as exemplified by the herpes viruses. The fast turnaround time in many of the new methods makes diagnostic microbiology much more clinically relevant for routine testing, whereas in the past, the long delay in culturing organisms in order to issue antibiotic sensitivity results made this testing less clinically relevant. The fast turnaround time of the new tests also results in rapid recognition of pathogens of public health significance such as multidrug-resistant bacteria, pandemic organisms and bioterrorism agents. More rapid turnaround time is also beneficial for infections which are rapidly progressive and dangerous, such as bloodstream infections.

Digital technology is based on the binary system and this has obviously produced quantum improvements in many facets of our daily lives. Nature, however, has provided us with a more sophisticated quaternary system – the four nucleotides in our genetic code. Microbiologists, as well as other life scientists, have exploited this system to develop and extend molecular tests for rapid detection, identification and confirmation of infection to the molecular level.

We would like to thank the authors and reviewers for their contributions to this complete electronic issue of *Microbiology Australia*.

Biographies

Tuckweng Kok is a virologist with 30 years' comprehensive experience in translational research, viral diagnostics and public health virology. He has been seconded to the WHO, Beijing as Virology Consultant for SARS, Human and Avian Influenza assessments. His research work focuses on neutralising antibodies to HIV and Influenza viruses and rapid viral and mycoplasma diagnostics. He has a passion for teaching and is one of the initiators of the postgraduate virology master class course at the University of Adelaide.

Iain Gosbell is a microbiologist and infectious diseases physician, and was previously the Director of the Department of Microbiology and Infectious Diseases, Sydney South West Pathology Service – Liverpool, and took up the position of Professor of Microbiology and Infectious Diseases in the School of Medicine, University of Western Sydney in late 2009. He has published on antibiotic-resistant bacteria, particularly community methicillin-resistant *Staphylococcus aureus* (MRSA). Recently in conjunction with Slade Jensen and Björn Espedido he helped co-found the Antibiotic Resistance and Mobile Elements Group (ARMEG) at the School of Medicine, which will examine resistance elements in multiresistant organisms (MROs), especially MRSA.

Rapid sequencing and analysis for pathogen confirmation

Rodney M Ratcliff

Viral Epidemiology, Microbiology and Infectious Diseases
Institute of Medical and Veterinary Science, SA Pathology
Adelaide, SA 5000
Microbiology and Immunology
School of Molecular & Biomedical Science, The University of Adelaide
Adelaide, SA 5005
Tel (08) 8222 3321
Fax (08) 8222 3543
Email
rodney.ratcliff@health.sa.gov.au

Genetic amplification methods and especially those based on the polymerase chain reaction (PCR) are revolutionising diagnostic microbiology. Genetic amplification is rapidly becoming the method of choice to detect viruses, 'single pathogen' infections such as Chlamydia and Gonorrhoea, slow growing or difficult-to-grow bacteria such as mycobacteria and legionella, or toxigenic bacteria such as Shigatoxigenic *Escherichia coli* and *Clostridium difficile*. This trend has been accelerated with the use of fluorescent, probe-based, real-time PCR, high-throughput thermocyclers and the use of robotics. Traditional biochemistry-based identification is adequate for many bacterial isolates, as is the specificity afforded by real-time probes for direct PCR detection. However, the sequencing of amplified products (amplicons) is also often performed where increased resolution is required, even though longer amplicons than those generated by real-time PCR are required. Two benefits are clear. Firstly, sequence is constructed solely from the four nucleotides so can be likened to a digital signal. In comparison, biochemical identifications are more like an analogue signal, where weak or equivocal reactions can create ambiguity. Secondly, the resolution afforded by biochemical reactions (less than 100 character states i.e. 48 reactions that can be positive or negative) is easily

exceeded by sequencing (2000 character states for a 500 bp sequence i.e. 500 sites which can contain one of four nucleotides). Further, the sequence is reproducible and not dependent on variables such as the bacterial growth conditions and easily lends itself to computer-based analysis.

Technical issues and strategies

Typically sequence is generated from PCR products using dye-terminator chemistry. The length of the product, which typically ranges from 100 to 1000 bases is determined by the availability of 'universal' primer binding sites and the level of sensitivity that is required. The former can be a challenge, especially for RNA viruses such as norovirus, where genomic variation among strains is high. This frequently necessitates the use of degenerate primers to achieve broad reactivity. Using degenerate primers lowers assay sensitivity so can be problematic if the target copy number is low, which can occur when direct detection and identification is being attempted. An unconfirmed rule of thumb is that sensitivity decreases 10-fold for every 100 bp increase in product length. Reverse-transcribing RNA targets, such as in the example above, further reduces sensitivity. Additionally, highly degenerate primers are often relatively poor at priming sequencing reactions even if they are adequate for the primary PCR detection. This problem can be solved by attaching an ~18-mer 5' tag to the amplification primer. M13-F and M13-R, sequence commonly incorporated in clones to permit sequencing the clone insert are the tags often chosen for the forward and reverse primers, respectively. The amplicon can then be sequenced with a primer specific for the tag only. Some laboratories choose this approach as a routine for all primers, so that all targets can be sequenced in both directions with just the two tag-specific primers, irrespective of the actual pathogen target. An alternative approach may be required where the choice of primers or cycling conditions required to achieve amplification is so restrictive that contaminating amplicons are also produced. Using an 'in-stepped' sequencing primer, with up to five 3' bases specific to the start of the legitimate product, or an internal primer, will generate unambiguous sequence in the presence of multiple amplification products.

Sequence quality

The quality of the sequencing electropherograms (the trace generated by the sequencer) is critical to the accuracy of all downstream interpretations. It is important to follow the manufacturer's methodology, especially the concentrations of target and primer, accurately to ensure clean, even, sigmoidal fluorescent dye peaks are obtained for the entire length of the sequence. To speed up the process, some laboratories have experimented with deleting one or both of the cleaning steps and reported acceptable sequence quality. Always review the electropherograms and if sub-optimal, work closely with the sequencing facility to ensure methodologies are optimised for the equipment. One complication to achieving quality sequence is simultaneously amplifying more than one related target. This can occur when there is more than one gene copy per cell, such as for rRNA gene targets (Figure 1) and for multiple variant strains of the same virus, such as HIV, in the patient's serum. Determining the exact sequence of each population if mixed bases or INDELs are present may require cloning, but this is rarely necessary.

Phylogeny

Phylogeny is the study of relationship between objects, typically the evolution of living cells. As discussed above, sequence variation is an ideal measure for this task. While the topic is

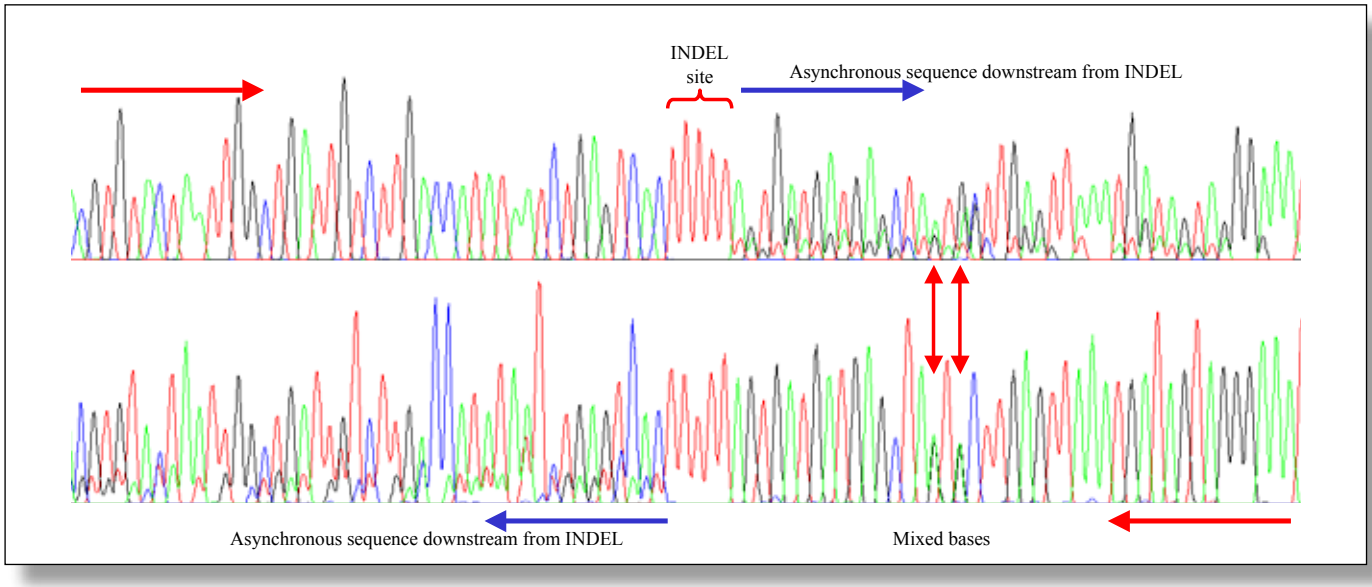

Figure 1. An alignment of forward and reverse sequence electropherograms from the 16S rRNA gene, demonstrating the effect of insertions or deletions (INDELS) and single nucleotide substitutions with respect to each other in multiple sequence populations. Note the loss of synchronicity immediately downstream from the INDEL. Two point mutations in one population are also highlighted, with mixed bases being evident in the sequence at that site.

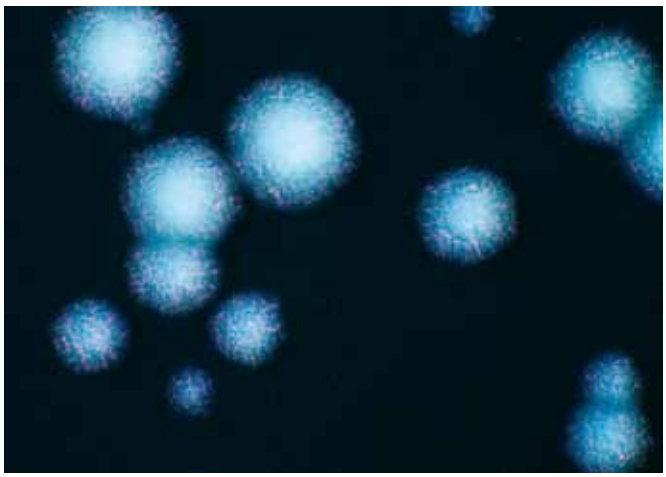

Figure 2. A *Legionella* strain growing on CYE agar. Note the typical opalescent ground-glass appearance.

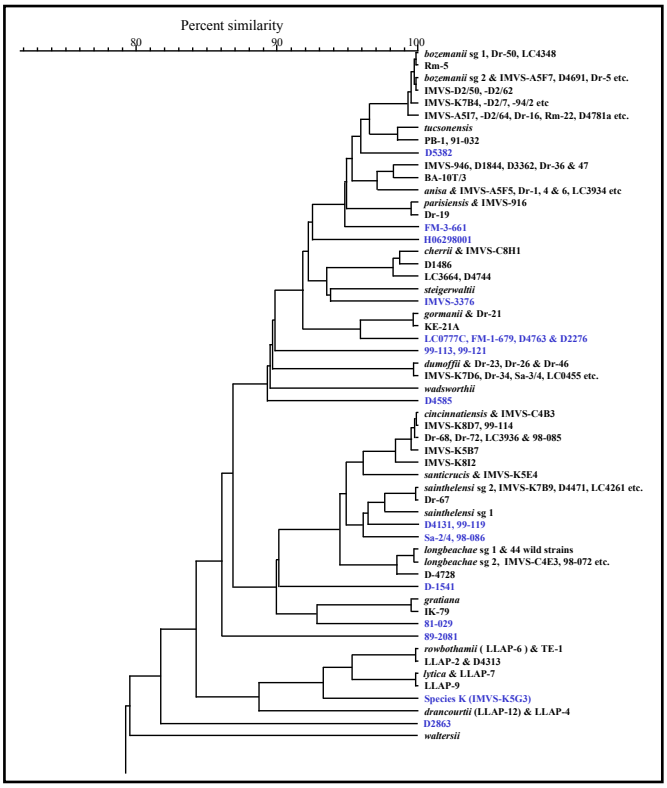

Figure 3. A section of a UPGMA similarity dendrogram of type and wild *Legionella* strains based on *mip* gene sequence. The complete dendrogram contains nearly 500 sequences. Wild strains have identical or similar sequence to that from the type strain of each species. *Legionella bozemanii* and *L. tucsonensis* are the most similar sister species (96.5%). Strains thought to be novel are highlighted in blue. This level of resolution can not be achieved using biochemical or serological profiles.

complex and a science in itself, sequence-based phylogenetic programs can quickly calculate the associations. Stand-alone programs, such as Kodon or Bioinformatics (Applied Maths, Belgium), provide a database approach where the sequences can be edited, stored and phylogenetically analysed locally. Alternatively, web-based systems such as BLAST permit the interrogation of large central databases such as GenBank. The results can be simply displayed as a list of nearest neighbour sequences (BLAST), or as an evolutionary tree (dendrogram), of which there are many formats, displaying the potentially complex associations between sequences graphically (for example, maximum likelihood, neighbour joining and so on). For the latter, only the nearest neighbour associations should be considered robust unless the analysis has been rigorous, incorporating the optimal evolutionary model, such as is permitted with the program PAUP (SinauerAssociates, Massachusetts).

Identification of *Legionella* isolates

Legionella strains are isolated from the culture of respiratory secretions or from environmental screening of water and soil samples (Figure 2). Legionellae are considered biochemically inert so the identification of isolates is difficult. Commercial agglutination kits are available for the serological identification of common species. However, the genus has over 50 species, half of which have been associated with human infection. As a result, sequence-based identification is the method of choice for all but the most frequently isolated species such as *L. pneumophila* and *L. longbeachae.* There are a number of gene targets now available, but that utilising approximately 650 bases of the macrophage infectivity potentiator *(mip)* gene is the most widely used [1] and an analysis database is accessible via the web (http://www.hpa.org.uk). The use of *mip* gene identification has identified an additional ~45 novel species (Figure 3). Further, a consensus sequence-based typing (SBT) scheme incorporating seven gene targets has also been developed to assist the epidemiological typing of clinical and environmental isolates of *Legionella pneumophila.* By linking sequence from this number of gene targets, the power to recognise strain identity has increased to that achieved by pulsed field gel electrophoresis (PFGE). In conjunction with EWGLINET (the European Working Group for *Legionella* Infections Network), the main purpose of this European scheme is to enable the linking of travel-associated strains that potentially have a common source and represent an outbreak, perhaps in an hotel or holiday resort, even though the strains may have been isolated from patients after they returned to their respective countries.

Norovirus outbreaks

Norovirus genogroup 2 is the most common pathogen associated with outbreaks of acute gastroenteritis. These occur mainly in residential care facilities and hospitals. The viral load in faeces and vomitus of infected persons can be very high and the infectious dose is low, so transmission is predominantly via person-to-person contact and aerosols. Sequence-based typing of strains over the past decade or so has revealed that infection commonly occurs with pandemic subgroup 4 strains (GII/4) that arise approximately every two years and spread globally via human travel [2]. The emergence of a new pandemic strain typically results in a surge in outbreaks. The strains are identified by the year in which they were first detected: 1996, 2002, 2004, 2006a and b (the latter predominantly causing outbreaks in Australia in 2007) and 2008. For strain typing purposes, a number of sentinel

laboratories around the world routinely sequence a region of ORF2, which encodes the major capsid protein. Increased outbreak activity or virus sequence variation that might signal the emergence of a new pandemic strain, which can then be communicated via the global network NoroNet as an alert for health authorities.

Genotyping of HIV

Patients infected with HIV are frequently on long-term, highly active antiretroviral therapy (HAART). These drugs typically act by binding to and blocking the function of viral proteins, inhibiting viral replication. The high frequency of mutation which occurs during HIV replication can result in therapy escape. This is partly countered by the concurrent use of multiple drugs. By monitoring the efficacy of therapy, when viral escape is detected by a rise in the patient's viral load, the patient's HIV virus can be sequenced to determine which mutations are accumulating, especially those that will cause an amino acid substitution that inhibits the binding of an antiretroviral drug. There is extensive, ongoing research into which mutations, either singly or in combination, will critically influence the efficacy of each antiretroviral drug. Within the laboratory, the genes encoding the proteins targeted by the patient's HAART, most frequently the Protease and Reverse Transcriptase enzymes, are amplified and sequenced. The sequence electropherograms are examined as mixed bases and occasionally INDELS can be present, arising from the various mutationally-derived HIV strains present in the patient. Once corrected, the interpretation of the predicted 'virtual' resistance is electronically performed using web-based drug resistance predicting algorithms available either commercially (e.g. Virco, http://www.vircolab.com) or publicly (e.g. Stanford University's HIVdb) to guide the choice of HAART drugs.

Conclusion

Using sequence for the information it contains is now common. The above examples detail a variety of such uses and demonstrate the increased resolution afforded by a sequence-based approach. The latest generation of sequencers which permit direct mass sequencing will only enhance the availability, resolution and thus the use of sequence-based data as a laboratory tool.

References

1. Ratcliff, R.M. *et al.* (1998) Sequence-based classification scheme for the genus *Legionella* targeting the *mip* gene. *J. Clin. Microbiol.* 36, 1560–1567.
2. Siebenga, J.J. *et al.* (2009) Norovirus illness is a global problem: emergence and spread of norovirus GII.4 variants, 2001–2007. *J Infect. Dis.* 200, 802–812.

Biography

Rod Ratcliff is a Principal Medical Scientist at the Institute of Medical and Veterinary Science, SA Pathology, and an Affiliate Senior Lecturer at the University of Adelaide. His diagnostic and research interests focus on molecular detection and typing of pathogens, with particular emphasis on viruses associated with acute gastroenteritis, HIV genotyping and *Legionella* taxonomy.

A pragmatic approach to screening for transmissible resistance in the *Enterobacteriaceae*

Justin Ellem & Jon Iredell

Centre for Infectious Diseases and Microbiology
Westmead Hospital, NSW

Determining the origin, spread and characteristics of important resistance genes and plasmids in Gram-negative bacteria may be more informative than determining the epidemiology of bacteria themselves. Infection control resistance surveillance and containment efforts should focus on the transmission characteristics of the resistance rather than those of the index organism at the point of recognition.

Detection of transmissible antibiotic resistance in Gram-negative bacteria

The predictable relationship between the antibiotic-resistant phenotype and the presence of the responsible gene in Gram-positive bacteria such as methicillin-resistant *Staphylococcus aureus* (MRSA) allows confident screening on phenotype or genotype in healthcare facilities. Gram-negative bacteria adapted to a polymicrobial environment such as the gut may have one or more of a large number of resistance mechanisms, many of which are interdependent and/or duplicative and some of which may be 'silent', that is they are not expressed or require cofactors to provide the resistance phenotype. Even those in which single-gene mechanisms are responsible may have any one of hundreds of different genes. These may be acquired from a number of different bacteria and their nomenclature and/or identification is often disorganised and confusing. However, new opportunities are now available with coincident advances in bioinformatics,