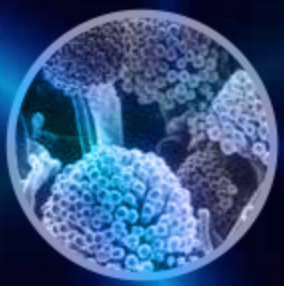
Molecular Methods in Plant Disease Diagnostics

Principles and Protocols



Edited by
Neil Boonham
Jenny Tomlinson
& Rick Mumford

Molecular Methods in Plant Disease Diagnostics

Molecular Methods in Plant Disease Diagnostics

Edited by

Neil Boonham

Fera, UK

Jenny Tomlinson

Fera, UK

and

Rick Mumford

Fera, UK

www.cabi.org

CABI is a trading name of CAB International

CABI
Nosworthy Way
Wallingford
Oxfordshire OX10 8DE
UK

Tel: +44 (0)1491 832111
Fax: +44 (0)1491 833508
E-mail: info@cabi.org
Website: www.cabi.org

CABI
745 Atlantic Avenue
8th Floor
Boston, MA 02111
USA

Tel: +1 (617)682-9015
E-mail: cabi-nao@cabi.org

A catalogue record for this book is available from the British Library, London, UK.

Library of Congress Cataloging-in-Publication Data

Names: Boonham, Neil, editor. | Tomlinson, Jenny, editor. | Mumford, Rick, editor. | C.A.B. International, issuing body.
Title: Molecular methods in plant disease diagnostics : principles and protocols / edited by Neil Boonham, Jenny Tomlinson, Rick Mumford.
Description: Wallingford, Oxfordshire ; Boston, MA : CABI, [2016] | Includes bibliographical references and index.
Identifiers: LCCN 2015036175 | ISBN 9781780641478 (alk. paper)
Subjects: LCSH: Plant diseases--Diagnosis. | MESH: Plant Diseases--genetics. | Molecular Diagnostic Techniques--methods.
Classification: LCC SB731 .M58 2016 | NLM SB 732.65 | DDC 632.3--dc23 LC record available at http://lccn.loc.gov/2015036175

ISBN-13: 978 1 78064 147 8

Commissioning editor: Alex Hollingsworth
Associate editor: Alexandra Lainsbury
Production editor: James Bishop

Typeset by SPi, Pondicherry, India
Printed and bound in Malta by Gutenberg Press Ltd, Tarxien, Malta

Contents

Editors and Contributors

Ian Adams, Fera, Sand Hutton, York, YO41 1LZ, UK. E-mail: ian.adams@fera.co.uk
Geraldine Anthoine, Plant Health Laboratory, Anses (French Agency for Food, Environmental and Occupational Health and Safety), 7 rue Jean Dixmeras, 49044 Angers cedex 01, France. E-mail: geraldine.anthoine@anses.fr
Neil Boonham, Fera, Sand Hutton, York, YO41 1LZ, UK. E-mail: neil.boonham@fera.co.uk
Andreas Bühlmann, Research Division Plant-Based Food, Agroscope, CH-8820 Wädenswil, Switzerland. E-mail: andreas.buehlmann@agroscope.admin.ch
Aude Chabirand, Plant Health Laboratory, Anses (French Agency for Food, Environmental and Occupational Health and Safety), Saint Pierre 97410, Réunion, France. E-mail: aude.chabirand@anses.fr
Catia Delmiglio, Plant Health and Environment Laboratory, Ministry for Primary Industries, 231 Morrin Road, St Johns, PO Box 2095, Auckland 1140, New Zealand. E-mail: catia.delmiglio@mpi.govt.nz
Tanja Dreo, National Institute of Biology, Večna pot 111, SI-1000 Ljubljana, Slovenia. E-mail: tanja.dreo@nib.si
Adrian Fox, Fera, Sand Hutton, York, YO41 1LZ, UK. E-mail: adrian.fox@fera.co.uk
Jürg E. Frey, Research Division Diagnostics and Risk Assessment Plant Protection, Agroscope CH-8820 Wädenswil, Switzerland. E-mail: juerg.frey@agroscope.admin.ch
Rachel Glover, Fera, Sand Hutton, York, YO41 1LZ, UK. E-mail: rachel.glover@fera.co.uk
Kristina Gruden, National Institute of Biology, Večna pot 111, SI-1000 Ljubljana, Slovenia. E-mail: kristina.gruden@nib.si
Catherine Harrison, Fera, Sand Hutton, York, YO41 1LZ, UK. E-mail: catherine.harrison@fera.co.uk
Jennifer Hodgetts, Fera, Sand Hutton, York, YO41 1LZ, UK. E-mail: jennifer.hodgetts@fera.co.uk
Tracey M. Immanuel, Plant Health and Environment Laboratory, Ministry for Primary Industries, 231 Morrin Road, St Johns, PO Box 2095, Auckland 1140, New Zealand. E-mail: tracey.immanuel@mpi.govt.nz

Subuhi Khan, Plant Health and Environment Laboratory, Ministry for Primary Industries, 231 Morrin Road, St Johns, PO Box 2095, Auckland 1140, New Zealand. E-mail: subuhi.khan@mpi.govt.nz

Sebastian Kiewnick, Plant Protection Division, Agroscope, CH-8820 Wädenswil, Switzerland. E-mail: sebastian.kiewnick@agroscope.admin.ch

Lynn Laurenson, Fera, Sand Hutton, York, YO41 1LZ, UK. E-mail: lynn.laurenson@fera.co.uk

Lia W. Liefting, Plant Health and Environment Laboratory, Ministry for Primary Industries, 231 Morrin Road, St Johns, PO Box 2095, Auckland 1140, New Zealand. E-mail: lia.liefting@mpi.govt.nz

Nataša Mehle, National Institute of Biology, Večna pot 111, SI-1000 Ljubljana, Slovenia. E-mail: natasa.mehle@nib.si

Rick Mumford, Fera, Sand Hutton, York, YO41 1LZ, UK. E-mail: rick.mumford@fera.co.uk

Maja Ravnikar, National Institute of Biology, Večna pot 111, SI-1000 Ljubljana, Slovenia. E-mail: maja.ravnikar@nib.si

Jenny Tomlinson, Fera, Sand Hutton, York, YO41 1LZ, UK. E-mail: jenny.tomlinson@fera.co.uk

Lisa I. Ward, Plant Health and Environment Laboratory, Ministry for Primary Industries, 231 Morrin Road, St Johns, PO Box 2095, Auckland 1140, New Zealand. E-mail: lisa.ward@mpi.govt.nz

Foreword

I joined the then Plant Pathology Laboratory of the Ministry of Agriculture in Harpenden in 1988. At that time, the diagnosis of plant pests and pathogens required mastering a very wide range of techniques from infection assays, electron microscopy, serology and culture methods through to morphological and taxonomic investigations. Each of these methods required many years of experience to perform and a great deal of specialization. One aim of our research programme then was to simplify these methods, while recognizing the ever-narrowing pool of expert taxonomists and other specialists that were either available or even being trained. We were looking for what we called 'generic' methods that could be applied across all taxonomic groups. Of course, the only common feature between, say, a viroid and a whitefly is its signature nucleic acid sequence, which offers a definitive and almost 'digital' result, irrespective of the life cycle or even the quality of the sample or specimen. So-called 'molecular' methods were born, and to this day I can remember back to 1990, when I was being instructed by my colleague, Christine Henry, on how to perform PCR tests for *Beet necrotic yellow vein virus*. I had two water baths, a heated block and an egg timer (and a good book to read).

Were we successful? My career since leaving Fera has taken me overseas and in particular to East Africa. During the last 2 years we have watched the development of a new and very serious disease of maize called 'Koroito' (the local language name for plague). At the same time, our seed potato programme was being affected by a potentially new millipede problem. The causal agent of the new maize disease was subsequently indicated by next-generation sequencing, and the seeds regulatory authority (the Kenya Plant Health Inspectorate Service, KEPHIS) now routinely tests seed samples using real-time PCR. I asked the taxonomists at the National Museums of Kenya about the identity of the new millipede pests we were encountering, and they replied that specimens had been sent for DNA barcoding. So I guess the answer must, pleasingly, be yes.

This book takes the practitioner through the full range of molecular diagnostic and detection methods and is authored by the leading people in the field. I highly recommend it to both beginners and experienced practitioners. From reading it, it is clear

that we have made great strides in molecular diagnostics since those first tentative steps back in the late 1980s/early 1990s. Is there still work to do? Yes, of course, challenges still remain. For the routine detection of plant pests and pathogens, when large numbers of samples need to be tested, then more can be done to bring down unit costs and make this technology more accessible. Further steps can be taken to permit staff in more modest laboratories to achieve high-quality and reproducible results routinely. But while these technical challenges remain and could perhaps form the subject of a future book, this current volume captures the state-of-the-art of a diagnostic discipline that has come of age in leaving the research bench and becoming established within the routine testing laboratory.

Ian Barker
Basel, 2015

1 Introduction: advances in plant diagnostics – historical perspectives and future directions

Rick Mumford*

Fera, Sand Hutton, York, UK

1.1 Progress in recent years

In 2006, the editors of this book co-authored a review entitled 'Advances in molecular phytodiagnostics – new solutions for old problems' (Mumford *et al.*, 2006). In this, they summarized the progress made during the first three decades of plant molecular diagnostics – since nucleic acid-based methods first started being used for the detection and identification of plant pathogens in the 1970s. In this introduction, we compare the state-of-the-art in the mid-2000s (as identified in the aforementioned review) with the current situation. This allows us to assess how this whole field of science has developed in the subsequent 9 years and where plant molecular diagnostics might be heading in the future.

1.2 The rise and rise of PCR

By 2006, the polymerase chain reaction (PCR) had become the dominant technology underpinning molecular plant diagnostics. Offering a whole raft of benefits including specificity, sensitivity, ease of assay design and applicability across a whole range of targets (pathogens and pests), PCR-based technology pushed other techniques such as dot-blot hybridization into the margins. By the mid-2000s, conventional PCR and, increasingly, real-time PCR were established within many diagnostic laboratories and were being used routinely. In subsequent years, this is a trend that has continued, especially the more wide-scale adoption of real-time PCR as a core diagnostic tool; despite having higher capital outlay and per-test costs, the advantages are sufficient to drive uptake, especially in larger, centralized laboratories. For example, if you look at national phytosanitary reference laboratories across Europe, all will now be equipped to perform

*E-mail: rick.mumford@fera.co.uk

methods based on this technology. This is also reflected in the number of standardized protocols that now feature real-time PCR methods, e.g. the EPPO Diagnostic Protocols (EPPO, 2010). Protocols for performing routine PCR-based tests, both conventional and real-time, are presented in this book (Chapters 2 and 3, respectively).

An essential part of the rise of PCR and real-time PCR and their adoption as core diagnostic technologies has been key developments that have occurred around the chemistry of nucleic acids (NAs), which had remained largely unchanged since it was first developed and introduced. Significant improvements have been made to the steps both pre- and post-PCR, with one such area being the development of better nucleic acid (NA) extraction methods. These have greatly enhanced both the quality and quantity of DNA and/or RNA obtained, with consequential improvements in test reliability. Advances have also improved the range of targets from which NAs can be effectively extracted, including difficult plant materials and other matrices, such as soil. This has greatly enhanced the range of ways in which PCR (and other molecular diagnostic methods) can be used, from the direct identification of isolated pests or pathogens through to the screening of bulk samples.

The other significant opportunity presented by PCR has been to provide the basis for DNA barcoding approaches. By combining broad-spectrum PCR primers (that amplify at taxonomic levels above species) with low-cost automated sequencing and better bioinformatics, routine sequencing-based identification of pests and pathogens is now both reliable and cost-effective. In 2015, this methodology has become established as a standard approach for both pest and pathogen identification, sitting alongside morphology-based methods and becoming increasingly essential where these traditional skills are either lacking or are unable to provide reliable taxonomic resolution; for example, in the identification of some juvenile life stages of pests. Methods for DNA barcoding both pathogens and pests are covered in this book in Chapters 5 and 6, respectively.

1.3 Quality matters

One of the major differences between 2006 and the present day has been the increased focus on molecular diagnostics as an entire process, from sampling through to the provision of results. As new techniques were developed, efforts were understandably aimed at improvements in the component steps that constituted the test itself – namely extraction, assay and analysis – and significant improvements were made. Yet as these tests became more readily accepted and entered routine usage, it became clear that basic measures of test performance (for example, defining a limit of detection and a test's cross-reactivity with related species) were not alone sufficient criteria for defining a robust test and delivering a consistent diagnostic performance. It is only in more recent years that terms such as repeatability and reproducibility have become standard words within the phytodiagnostician's vocabulary. In order to overcome greater legal and regulatory scrutiny and challenge, laboratories have to provide greater quality assurance on the testing they carry out. This need for comparability is driving the development of international diagnostic protocols, both at a regional (e.g. EPPO, the European and Mediterranean Plant Protection Organization) and a global (e.g. IPPC, the International

Plant Protection Convention) level. Internationally recognized quality standards such as ISO 17025 are also being adopted. As a requisite of this, laboratories are now having to invest in additional physical infrastructure and processes (for example, to prevent cross-contamination), alongside the need for better controls (such as certified reference materials) and additional validation studies. As a result, guidance on quality assurance and diagnostic processes is required to run a modern molecular diagnostics laboratory (see Chapter 10 of this book).

1.4 The move from laboratory to field

A decade ago, plant pathologists had begun to evaluate platforms that allowed real-time PCR to be performed in the field. Projects such as *PortCheck* took technologies such as SmartCycler (Cepheid, USA) and developed simplified protocols, adapted from those used in the laboratory, that ran reliably and permitted real-time PCR to be performed at sites away from centralized diagnostic laboratories for the first time. Indeed, by the late 2000s, the UK's Plant Health and Seeds Inspectorate (PHSI) was using this technology at remote sites in the south-west of England as part of its *Phytophthora ramorum* eradication campaign.

However, despite this progress, it became apparent that PCR-based methods had significant limitations as a field-deployable technology, and the search for the next-generation of on-site DNA testing began. A significant part of that investigation was to identify efficient isothermal chemistries that could act as effective alternatives to PCR. Of the many alternatives, loop-mediated isothermal amplification (LAMP) has emerged as the current preferred choice among plant diagnosticians and the chemistry that is being adopted for on-site testing. Together with the use of different chemistry, further advances have taken place based on simplifying extraction and developing better workflows and detection systems. In combination, these advances have made routine on-site molecular diagnostics a reality. A decade on from the first papers published that described the use of LAMP for plant pathogen detection, technology based on this chemistry is now being deployed with plant health inspectors and will start to become an integral part of quarantine monitoring and surveillance across Europe. Methods describing LAMP (Chapter 4) and on-site testing approaches (Chapter 9) are included in this book.

1.5 Identifying the unknown

While the methods described above offer solutions for specific target detection, in the laboratory or in the field, and approaches for more accurate identification, they all assume a level of understanding of what target is actually being tested. In many cases, the real challenge for diagnosticians is 'unknowns'. These might be new diseases, for example, or known pathogens on a previously unrecorded host, complex disease syndromes where the causal agent is unclear or new pests and pathogens that have not been previously characterized. A classic example of this is blackcurrant reversion, where the virus causing the disease (*Blackcurrant reversion virus*, BRV) took decades to isolate and identify. In all of these scenarios, traditional approaches require multiple parallel

tests to be performed, often based on a range of different techniques. This can be slow and frequently unsuccessful, e.g. where it proves impossible to isolate the causal agent. This challenge has led diagnosticians to investigate multiplex platforms that are capable of simultaneously detecting multiple targets in a single test. As described in a review by Boonham *et al.* (2007), microarray technology had become the main focus for this approach. Starting initially with viruses and glass slide arrays, over time this technology has developed to cover not only additional target taxa but also alternative formats such as tube arrays. An account of microarray detection procedures and protocols is provided in Chapter 7 of this book.

However, while arrays offer the potential to test for hundreds, if not thousands, of pathogens in a single test, they use specific probes designed against known targets that have been previously identified and for which published genome data exist. Hence, while microarrays can help to identify known pathogens causing new diseases, they will still fail to identify new or highly distinct strains of pathogens. In 2005–2006, the solution to this problem remained elusive. Yet it was at that time that the first next-generation sequencing (NGS) platforms became available. With their huge *de novo* sequencing capacity, for the first time these new platforms offered a non-targeted means to generate sequences from the entire DNA complement present in a sample. Using advanced bioinformatics approaches, it would be possible to analyse the huge data sets generated and specifically identify those known or putative pathogen sequences (Studholme *et al.*, 2011). By the late 2000s, this approach had been applied successfully for the detection of novel plant viruses. Further developments, new platforms, better bioinformatics and decreasing per run sequencing costs have seen this technology start to become adopted as a frontline diagnostic tool. As the technology progresses, this is a trend that will only continue to gain momentum. NGS diagnostic approaches and methods are covered in this book in Chapter 8.

1.6 Future challenges

In this short introduction, it has been possible to describe the huge progress made in plant molecular diagnostics since the 1970s and 1980s, when these methods first started to be used. The introduction has also highlighted the rapid progression made during the last decade or so in turning these methods from being specialist applications, carried out by experts in centralized laboratories, into routine methods that are everyday tools for diagnosticians and, increasingly, for non-specialists. The publication of this book provides clear evidence of that. None the less, for all of the progress made, challenges still remain and these will need to be addressed in future years if molecular diagnostics are to become even more widely adopted and the benefits offered by them fully realized (Boonham *et al.*, 2014).

The first major challenge is around sampling and nucleic acid extraction. As already noted, significant progress has been made in terms of moving to quicker and more reliable methods that avoid the use of noxious solvents such as phenol. Yet in the 21st century, we still rely predominantly on one single extraction chemistry, namely binding to silica in the presence of a chaotropic agent (the 'Boom' method). While different formats now exist, including columns and magnetic particles, and provide greater flexibility for automation,

this extraction chemistry also presents limitations is terms of capacity and the ability to deal with inhibitors. New chemistries and innovative platforms could certainly create new opportunities for innovation in diagnostics. Current methods for the preparation of plant samples through grinding, milling or pulverization present a major bottleneck for high-throughput testing. The development of better sample preparation equipment, which combines a low per sample cost with reduced labour input and minimal cross-contamination risk, would be a panacea for many diagnosticians. Going a step further back, the process of sampling prior to extraction continues to be a stage that is exclusively manual and very labour intensive. The development of better dedicated sampling devices would help to accelerate the testing process, aid consistency and benefit the staff carrying out this work. In the longer term, the integration of sampling, extraction and testing within a single platform or device is the ultimate goal for many working in this field. However, this will only be achieved through the implementation of a truly interdisciplinary science approach, combining the expertise of biologists, chemists and engineers alike.

Another key area where future efforts need to be focused is the development of integrated, quality-assured diagnostic workflows that cover the entire diagnostic process. Alongside the obvious advances in the testing methodology, there will need to be further developments in the diagnostic support services that work with these new technologies. Key examples would be the development of better proficiency testing schemes for plant pests and pathogens – ones that are not only applicable to laboratory testing but also to field testing. With new technologies such as NGS, new approaches will be required to verify not only the generation of sequence data sets but also their analysis – the design will be needed of new ways to verify and approve bioinformatics pipelines. As sequence-based identification becomes increasingly important, more effort will be required to establish verified databases of voucher sequences that have been confirmed as belonging to the species in question. This must build on approaches already being developed under initiatives such as the European initiative Q-Bank (a collection of comprehensive databases on quarantine plant pests and diseases available at www.q-bank.eu). Finally, as we see the expanding devolution of diagnostics from the laboratory to the field, this further strengthens the need for expert reference laboratories with the ability not only to verify and confirm field diagnosis, but also to support the performance of field testing with advice, training and other diagnostic services, e.g. proficiency testing. These reference laboratories will in addition provide an essential role in the identification of new and emerging pests, a role which is unlikely to be easily achieved in the field using rapid, targeted diagnostics.

The final big challenge is data and data sharing. With the rise of new technologies such as NGS, diagnosticians will face new challenges in how to handle, analyse and store the huge data sets generated, which dwarf those generated by previous techniques. There will also be challenges linked to sharing data between field and laboratory as diagnostics become more portable. There is a drive to encourage better data sharing between organizations, both nationally and internationally, and this will require the adoption of standard data formats. All of these drivers will mean a much greater role for bioinformaticians, software designers and other knowledge management professionals working along with diagnosticians. So, as described earlier, this is another area of plant diagnostics where interdisciplinarity will need to become a major theme in the future.

References

Boonham, N., Tomlinson, J. and Mumford, R.A. (2007) Microarrays for rapid identification of plant viruses. *Annual Review of Phytopathology* 45, 307–328.

Boonham, N., Kreuze, J., Winter, S., van der Vlugt, R., Bergervoet, J., Tomlinson, J. and Mumford R. (2014) Methods in virus diagnostics: from ELISA to next generation sequencing. *Virus Research* 186, 20–31.

EPPO (2010) Standard PM 7/76: use of EPPO diagnostic protocols. *EPPO Bulletin* 40, 350–352.

Mumford, R.A., Tomlinson, J., Barker, I. and Boonham, N. (2006) Advances in molecular phytodiagnostics – new solutions for old problems. *European Journal of Plant Pathology* 116, 1–19.

Studholme, D.J., Glover, R.H. and Boonham, N. (2011) Application of high-throughput DNA sequencing in phytopathology. *Annual Review of Phytopathology* 49, 87–105.

2 Conventional PCR

Lisa I. Ward,* Tracey M. Immanuel, Subuhi Khan, Lia W. Liefting and Catia Delmiglio

Plant Health and Environment Laboratory, Ministry for Primary Industries, Auckland, New Zealand

2.1 Introduction

The discovery of DNA by James Watson and Francis Crick in 1953 founded the field of molecular biology (Watson and Crick, 1953); it took another 30 years of technological development, including the making of synthetic DNA (oligonucleotides) and the discovery of heat-tolerant enzymes, to lead to the development of the polymerase chain reaction (PCR) in 1983 (Shampo and Kyle, 2002).

The principle of PCR is to make millions of copies of a target DNA molecule in a very short time. The technique was originally applied to DNA, but has since been adapted for the amplification of RNA. The simplicity, repeatability and speed of the PCR technique has revolutionized molecular biology and allowed, among other things, the study of human evolution, the diagnosis of genetic diseases, and the development of DNA fingerprinting and the ability to detect pathogens in humans, animals and plants.

PCR uses a mixture of heat-tolerant enzymes (DNA polymerase), oligonucleotides (primers) and nucleotides (nts: adenine, A; thymine, T; cytosine, C; and guanine, G) to copy a small fragment of a target DNA in a sequence-specific fashion. As the name suggests, the biochemical reaction is put through a cycle of heating to melt the double-stranded (ds) DNA, cooling to allow the primers to bind (anneal) to the target DNA and further heating to allow the polymerase enzyme to copy the target DNA strand. This thermal

*Corresponding author: lisa.ward@mpi.govt.nz

cycling process allows the DNA template to be exponentially amplified until the reaction components have been depleted and the reaction plateaus. The amplified DNA can be stained with an intercalating dye and visualized following size separation on an agarose gel.

Since the development of PCR in the 1980s, there have been a number of modifications to the standard protocol that have enabled the development of a wide range of different techniques, such as **multiplex PCR**, in which primers for more than one target can be used together in the same reaction, and **hot start PCR**, in which the reaction components are heated to the DNA melting temperature (e.g. 95ºC) before adding the polymerase – in this way, non-specific amplification caused by primer binding at lower temperatures is prevented. **Touchdown PCR** is where the primer annealing temperature is gradually decreased in later cycles; the initial higher annealing temperature leads to greater specificity for primer binding, while the lower temperatures in later cycles permit more efficient amplification at the end of the reaction. **Nested PCR** uses two sets of primers in two successive reactions; the PCR products from the first reaction act as a template for the second reaction, thus increasing specificity and sensitivity. **Assembly PCR** allows the synthesis of DNA structures by performing PCR on a pool of long oligonucleotides with short overlapping segments in order to assemble two or more pieces of DNA into one piece (Stemmer *et al.*, 1995).

In the field of plant pathology, there has been an avalanche of newer technologies over the last 10–20 years, including quantitative (real-time) PCR (Boonham *et al.*, 2004; Harper *et al.*, 2011), isothermal amplification technologies (Haible *et al.*, 2006; Tomlinson *et al.*, 2007; Harper *et al.*, 2010), DNA microarrays (Boonham *et al.*, 2007) and, more recently, metagenome sequencing (Al Rwahnih *et al.*, 2009; Coetzee *et al.*, 2010). Despite these developments, the relatively low cost of conventional PCR, coupled with its versatility and speed, means it has remained a useful diagnostic tool in plant pathology. Some of the applications of conventional PCR for plant pathogens are discussed in this chapter.

2.2 Viruses and viroids

Conventional PCR remains an important technique for virology diagnosticians. One important approach is the use of generic or universal primers. Viruses from the same family or genus can have regions of high sequence similarity, and generic primers can be designed for these regions. This allows the detection of a number of species of viruses in a single PCR reaction. When designing generic primers that are specific for a virus genus or family group, it is important to make alignments of sequence from as many isolates as possible, and these can be obtained from GenBank (the US National Institutes of Health (NIH) genetic sequence database available from NCBI, the US National Center for Biotechnology Information). Designing primers using multiple isolates will make sure that the primers cover variant sequences; otherwise, there is a risk that some isolates will not be detected. It is also important to check the homology of primers to other sequences in GenBank to ensure that the primers are only detecting the virus group for which the primers are designed; this will avoid the amplification of non-target viruses. Primer homology can be checked against all available sequences in GenBank using the Basic Local Alignment Search Tool (BLAST). Generic primers for the virus groups *Carlavirus*, *Carmovirus*, *Dianthovirus*, *Ilarvirus*, *Tobamovirus* and *Tospovirus* are commercially available (from Agdia).

PCR products amplified using generic primers can be sequenced to enable identification of the virus species amplified. It is important to consider that more than one virus from a group may be present in the sample; this is sometimes evident from the presence of multiple peaks in a Sanger sequence chromatogram, which indicate multiple overlapping sequences. To identify the multiple viruses in the sample, the PCR amplicon can be cloned into a vector such as pTOPO® (ThermoFisher Scientific) and then sequencing multiple individual clones to determine which viruses are present. However, one limitation of generic primers is that they can be less sensitive in detecting viruses than specific primers.

For viroids, the presence of conserved regions enables amplification of the complete small genome (<400 nts) using conventional PCR (e.g. Puchta and Sanger, 1989; Ward *et al.*, 2011). Entire genome sequences of viroids can be easily compared by alignment and variants easily identified. Compared with the less specific and more laborious techniques such as PAGE, PCR has made viroid diagnostics much easier, with huge savings in detection time and cost.

2.3 Phytoplasmas

Phytoplasmas cannot be cultured *in vitro* and are therefore classified using the '*Candidatus*' concept in which each of the major clades established by 16S rRNA gene sequence analysis represents a *Candidatus* species of the *Phytoplasma* genus (IRPCM Phytoplasma/Spiroplasma Working Team – Phytoplasma Taxonomy Group, 2004). Because of the highly conserved nature of the 16S rRNA gene, at least a 1.2 kb fragment needs to be amplified and sequenced from the phytoplasma 16S rRNA gene in order to determine its identity. This 1.2 kb fragment of the 16S rRNA gene is amplified using the R16F2n (Gundersen and Lee, 1996) and R16R2 (Lee *et al.*, 1993) primers. Phytoplasmas can be present at a low titre in their plant host and therefore quite often can only be detected by using nested PCR or real-time PCR. For nested PCR, the P1 (Deng and Hiruki, 1991) and P7 (Schneider *et al.*, 1995) primers are used in the first-stage PCR and the resulting PCR products are re-amplified using the R16F2n and R16R2 primers. These two sets of PCR primers are universal and will detect all known phytoplasmas. False positives with closely related bacteria can be obtained using PCR primers designed to the phytoplasma 16S rRNA gene. Therefore, if the outcome is critical, the PCR product should be sequenced.

PCR methods have also been used to amplify other genome regions for phytoplasma detection and classification, including ribosomal protein genes (Lim and Sears, 1992; Lee *et al.*, 1998; Martini *et al.*, 2007), the *tuf* gene (Makarova *et al.*, 2012), the 23S rRNA gene (Guo *et al.*, 2003; Hodgetts *et al.*, 2008), the *secA* gene (Hodgetts *et al.*, 2008) and the *secY* gene (Lee *et al.*, 2010). These PCR primers are group specific and do not amplify DNA from phytoplasmas in other groups. However, they may be useful when a second independent region of the phytoplasma genome is required or for more detailed subgroup delineation. To date, universal amplification of all phytoplasmas can only be achieved with primers designed to the 16S rRNA or 23S rRNA genes. Attempts to design universal phytoplasma primers from other genes have not been successful. Although the primers designed by Hodgetts *et al.* (2008) to the *secA* gene seemed promising, they

were only tested on 34 phytoplasma strains and they failed to amplify additional strains (Makarova *et al.*, 2012). Most recently, Makarova *et al.* (2012) evaluated a nested PCR assay based on the *tuf* gene for DNA barcoding of phytoplasmas. In order to amplify all 91 phytoplasma strains tested, a primer cocktail was used: five primers in the first-stage PCR and nine in the second-stage PCR. This *tuf* gene DNA barcoding assay failed to amplify two phytoplasma strains in our laboratory and instead amplified the *tuf* gene from the plant host or another bacterium.

2.4 Bacterial pathogens

PCR can be used to overcome some of the limitations of traditional microbiological techniques, such as plating and biochemical methods, for the detection of bacterial pathogens. The challenges of plating methods (selective or non-selective) include the inability to detect non-culturable or fastidious bacteria and the rapid growth of saprophytes – which can swamp colonies of the pathogen and may prevent detection. PCR can circumvent many of the above drawbacks and can be used to test plants directly, in some cases alleviating the need for plating altogether.

The development of either specific or generic primers for bacterial organisms relies on good molecular taxonomic foundations. For example, the *Pseudomonas syringae* species group has been intensively studied, and a large amount of DNA sequence material from type strains exists on the core and flexible genomes of this group (Sarkar and Guttman, 2004). Universal primers have been designed for species and genus discrimination in the rDNA operon and include the internal transcribed spacer region (ITS) and the 16S and 23S rRNA genes (Li and De Boer, 1995; Maes *et al.*, 1996; Boyer *et al.*, 2001). Sequences obtained by using any of the rDNA-based primers can be uploaded to the Ribosomal Database Project (RDP; available at http://rdp.cme.msu.edu/), which enables them to be aligned against a repository of annotated rRNA sequences; the alignment can be downloaded to aid phylogenetic identification. The disadvantages of these universal primers include the need to isolate a culture first to avoid the co-amplification of environmental or saprophytic bacteria that may be associated with the host. The use of these universal primers can also lead to incorrect bacterial identification due to a lack of unique rRNA sequences for a particular bacterium in publicly available databases, or to insufficient sequence diversity among different bacterial species.

From a quarantine perspective, one of the biggest diagnostic challenges is the detection of new and emerging pathogens. A number of different methods of PCR have been developed to allow the typing and discrimination of genetic differences between populations. These include fingerprinting methods such as random amplified polymorphic DNA (RAPD) (Măruţescu *et al.*, 2009; Puławska and Sobiczewski, 2012), repetitive sequence based-PCR (rep-PCR) (Scortichini *et al.*, 1998; Puławska and Sobiczewski, 2012), amplified fragment length polymorphism (AFLP) (Avrova *et al.*, 2002; Pitman *et al.*, 2010) and restriction fragment length polymorphism (RFLP) (Manceau and Horvais, 1997). Multilocus sequence analysis (MLSA) of housekeeping genes and rapidly evolving effector and phytotoxin genes can also be used to elucidate the phylogenetic relationships between geographical isolates. MLSA was recently used to compare isolates of

Pseudomonas syringae pv. *actinidiae* (*Psa*) from around the world (Chapman *et al.*, 2012). *Psa* causes bacterial canker of kiwifruit (*Actinidia* spp.) and is an emerging pathogen that has shown an increase in disease incidence and geographic range (Taylor *et al.*, 2014). MLSA was successfully used to show the evolutionary trajectories, and hence the phylogenetic groups, of this pathogen without whole genome sequencing.

The known pathogenicity of some bacterial species or their phytotoxin genes have been used for designing primers to enable their specific detection. Examples include the use of the *nec1* gene for the detection of *Streptomyces* (Cullen and Lees, 2007), the phaseolotoxin gene for the detection of *Pseudomonas savastanoi* pv. *phaseolicola* (Oguiza *et al.*, 2004), the *pel* gene for detection of pectolytic species of the soft rot causing *Pectobacterium* spp. (Moleleki *et al.*, 2013) and the *hop*, *hrp* and *avr* genes for the detection of *Psa* (Chapman *et al.*, 2012). The disadvantage of using these genes for PCR detection is that pathogenicity genes can sometimes be missing even in pathogenic strains, thus suggesting that they play a secondary role in pathogenicity (Wanner, 2004, 2006).

The success of detecting bacterial pathogens by PCR may be influenced by the number of copies of the target DNA sequence. Enrichment methods can be used to increase the target bacterial numbers, and these include growing a culture in liquid or solid media prior to DNA extraction; this type of method has been shown to be useful for some matrices, such as seed and soil, or for latent infections of plants (López *et al.*, 2003). The quality of DNA extracts can be improved by using one of the many commercial kits available. Kits are available both for pure cultures and for direct extraction of bacterial DNA from plants. Two examples of well-tested kits for plant pathogenic bacteria are Qiagen's DNeasy® Plant Mini Kit and DNeasy® Blood and Tissue Kit. Cruder methods of extraction include fast prep kits such as Sigma's Extract-N-Amp Plant Kit. This latter kit enables DNA extraction in solution without the need for the freezing of plant tissues with liquid nitrogen, mechanical disruption, organic extraction, column DNA purification or alcohol precipitation. The resulting DNA is suitable for use in hot-start PCR, for which ready mixes are also available. An even cruder form of DNA extraction involves heating bacterial cells in water at high temperature to rupture the cells and using the resulting DNA directly in PCR; this technique is more suited to Gram-negative bacteria which do not have a thick peptidoglycan layer in their cell walls, therefore allowing easier cell wall disruption.

It is important when designing specific PCR assays for plant pathogenic bacteria, whether for species- or genus-level detection, that they are well validated by testing against a large number of isolates and strains of the target bacterium. They also need to be tested for cross-reaction against other bacterial pathogens and common saprophytes that occur on the same host. Alvarez (2004) stressed the importance of international ring tests among laboratories to test robustness, repeatability, ease of implementation and interpretation of test results in the routine analysis of large numbers of samples. Ring testing is especially important to build confidence in new PCR assays that have been developed for high-impact pests that are of international significance. Examples of PCR protocols that have been developed for quarantine bacteria and validated by ring tests include those for *Clavibacter michiganensis* subsp. *sepedonicus*, *Xanthomonas*

fragariae and *Erwinia amylovora*. Following ring testing, these PCR assays were taken up as recommended assays in EPPO (European and Mediterranean Plant Protection Organization) diagnostic protocols (Palacio-Bielsa *et al.*, 2009).

2.5 PCR detection of fungi

The molecular diagnostics of fungal pathogens is becomingly increasingly important, particularly in relation to detecting pathogens of international and quarantine significance. Traditional methods of detection involve culturing pathogens *in vitro*, and identification using morphological characteristics such as fruiting bodies, spores, hyphae and colony shape and colour. The identification process can be time-consuming and requires technical expertise in taxonomy.

PCR primers for fungal diagnostics can be designed for either conserved or variable regions. The most commonly used regions for PCR are the ITS regions (ITS1 and ITS2) of the fungal nuclear ribosomal DNA (rDNA); there can be up to 200 copies in a haploid genome and this high copy number can improve the sensitivity of detection by PCR. The ITS region has been accepted by the Consortium for the Barcode of Life (CBOL) as the official DNA barcode genetic marker for fungi, as it can be used to discriminate the broadest range of fungi to species level (Schoch *et al.*, 2012). Once amplified by PCR, the ITS region can be sequenced and the sequences compared with others in publicly available sequence databases such as GenBank or UNITE (the 'Unified system for the DNA based fungal species linked to the classification'). Difficulties in species identification may be encountered if there are only small sequence differences between and within a species, and ITS sequences cannot usually discriminate between taxa below the species level.

Although the ITS1 region is the most commonly used target for diagnostic PCR, other genes are increasingly being studied as alternative targets. Within the rDNA, the intergenic spacer sequence (IGS) region between the 28S and 18S rRNA genes is sometimes used when ITS regions are found to be too variable (Validov *et al.*, 2011).

The β-tubulin nuclear housekeeping gene is being increasingly used for fungal diagnostics, and for deep-level phylogenies and studies of species complexes (Fraaije *et al.*, 1999; Hirsch *et al.*, 2000; Einax and Voight, 2003). β-tubulin is most abundant in eukaryotic cells and its heterodimers are the primary constituent in the microtubules of fungi, which makes it suitable for the development of sensitive PCR assays. Other genes mentioned in the literature as being used for diagnostics and phylogeny include mating type genes (Foster *et al.*, 2002), the multicopy mitochondrial genes *cox I* and *cox II* (Martin *et al.*, 2009), the genes for translation elongation factor 1 alpha (EF1α) (Geiser *et al.*, 2004) and calmodulin (Mulè *et al.*, 2004), and avirulence genes (Lievens *et al.*, 2009). To enhance the specificity of a diagnostic assay, a multiple-multilocus approach may give greater certainty in fungal identification. Examples of this approach can be found in the literature (Collado-Romero *et al.*, 2008; Dixon *et al.*, 2009; Glienke *et al.*, 2011).

If species-specific primers are being used for PCR, then DNA can be extracted directly from infected plants. The disadvantage of testing infected plants directly is that it is hard to determine whether the target fungus is living or dead. To get around this, mRNA can

be extracted and tested by reverse-transcription PCR (RT-PCR); mRNA is degraded rapidly in dead cells and so its detection by RT–PCR is potentially an accurate indicator of cell viability (Sheridan *et al.*, 1998; Chimento *et al.*, 2012).

Universal primers would not normally be used to test infected plant material directly owing to the presence of DNA from other saprophytic fungi that could be co-amplified in PCR. In view of this, pure cultures of fungal pathogens should be isolated, followed by PCR testing. In order to do this, the infected plant tissue needs to be surface sterilized (e.g. with 1% sodium hypochlorite). Samples can be taken at the junction of diseased and healthy tissue and these are then plated on either non-selective or selective media. Alternatively, the fungus can be encouraged to sporulate on plant tissues in a moist chamber and under UV light, followed by the transfer of single spores on to media. Once a pure culture is obtained, this can be tested directly by PCR, usually following DNA isolation.

2.6 Internal controls

As in any diagnostic, it is important to use appropriate controls for the most critical steps of the process. The incorporation of internal amplification controls, often designed with plant genes, helps to exclude false negatives caused by inhibiting compounds or failed nucleic acid extraction (Weller *et al.*, 2000). For the detection of fungal, bacterial and phytoplasma targets, controls based on conserved genes such as the cytochrome oxidase (COX) gene are often selected, enabling one assay to be used across many host species. For viruses, an RNA-only control may be appropriate and the mitochondrial gene *Nad5* (for NADH dehydrogenase subunit 5) is commonly used to check the competency of plant RNA extractions. The *Nad5* gene contains two exons separated by an intron. Kato *et al.* (1995) designed a *Nad5* sense primer with the first 20 nts homologous to the first exon and the last three nts homologous to the second exon. Thus, the primer covers the exon–intron boundary, which ensures that mRNA is being detected rather than DNA. As the majority of plant viruses have an RNA genome, using these primers (Kato *et al.*, 1995; Menzel *et al.*, 2002) to detect *Nad5* mRNA ensures that the RNA is of sufficient quality to amplify an RNA virus if it is present.

2.7 PCR optimization

An optimized PCR will improve the specificity, sensitivity and robustness of the reaction, while reducing the amplification of non-specific products. Some of the parameters that need to be considered for PCR optimization are discussed in the following sections.

2.7.1 Primers

While primer design software makes the design process simple, frequently calculated primer pairs do not work in practice and optimization is required. The first step for primer design is to find regions of sequence that are specific for the target but which should not amplify other genes or non-target organisms. The number of primer homologies within a complex genome is one of the most frequent causes of PCR failure (Andreson *et al.*, 2008). The BLAST tool (Altschul *et al.*, 1990) and, more commonly, BLASTn (nucleotide BLAST),

have been widely used to assess the number of homologies. Using BLAST against the appropriate non-redundant database helps to avoid regions of cross homology that can cause amplification of sequence from non-target organisms (see the primer design guidelines available at http://www.premierbiosoft.com/tech_notes/PCR_Primer_Design.html). Other important parameters to be considered while designing primers are: melting temperature (52–65°C), GC content (40–60%), length (18–22 bp), potential for the formation of hairpins and heterodimers and homodimers, and avoiding cross-reaction to non-target templates (see primer design tips and tools available at http://www.thermofisher.com/uk/en/home/products-and-services/product-types/primers-oligos-nucleotides/invitrogen-custom-dna-oligos/primer-design-tools.html). The calculated melting temperature (T_m) is especially important for PCR as it determines the optimal annealing temperature of the primer in the reaction. Primer concentrations should be high enough to enable sufficient sensitivity, specificity and fidelity within the assay. However, using high primer concentrations does increase the risk of primer–dimer formation. The optimal concentrations for a PCR reaction are calculated based on testing, but a range between 200 and 600 nM is recommended as optimal. Another variable to look at is the inclusion of a G or C residue at the 3′ end of primers. This 'GC Clamp' helps to ensure correct binding at the 3′ end owing to the stronger hydrogen bonding of G–C residues.

2.7.2 Cycling conditions

Even with careful primer design, thermal cycling conditions should be optimized, in particular the annealing temperature, which is best achieved by setting up an annealing gradient of ±10°C of the calculated primer T_m (Ishii and Fukui, 2001; Roux, 1995). The best primer sets are those that give single amplification products over the widest annealing temperature range (Crews *et al.*, 2008). Although the annealing temperature is probably the single most important parameter that affects sensitivity (Roux, 1995), denaturation temperatures can also affect the PCR result. For example, too high denaturation temperatures and too long denaturation damage cytosine residues, leading to deamination, can also damage the DNA polymerase, thus reducing PCR performance (Wittwer and Garling, 1991). A lower denaturation temperature can be used to aid the amplification of longer or more difficult templates. The PCR extension time and temperature depends on the DNA polymerase and the template, but generally 1 min per kilobase (kb) product length and a temperature of 65–72°C is recommended. Alternatively, other PCR protocols are adopted when standard PCR protocols do not achieve the required results, such as touchdown PCR, in which the annealing temperature is gradually decreased in later cycles (Don *et al.*, 1991) so that it is more stringent in early PCR cycles, thereby reducing the amplification of undesired products (Hecker and Roux, 1996). In addition to touchdown PCR, some of the other variants that can be used in this situation are nested PCR, multiplex PCR, RT-PCR, assembly PCR and asymmetric PCR (Warrens *et al.*, 1997).

2.7.3 Reaction components

Using optimal concentrations of common reagents such as magnesium, deoxynucleotide triphosphates (dNTPs) or DNA polymerase can improve PCR results.

Magnesium titrations are commonly performed as magnesium ions act as a cofactor for thermostable DNA polymerases (e.g. *Taq* polymerase) and fluctuations in concentrations can greatly affect the enzyme performance. Too much magnesium can inhibit the PCR by stabilizing ds DNA and preventing the complete target denaturation, so increasing the error rate of the DNA polymerase (Wiedbrauk *et al.*, 1995; Markoulatos *et al.*, 2002). Optimal magnesium concentrations range from 1 to 6 mM and the final concentrations vary significantly between commercial master mixes (Wiedbrauk *et al.*, 1995). The concentration of dNTPs is also a limiting factor in PCR and an imbalance of dNTPs may reduce the overall fidelity of the enzyme (Kunz and Kohalmi, 1991). Typical concentrations range from 50 to 2000 µM, with long PCRs requiring more dNTPs. An increase in dNTPs requires a commensurate increase in magnesium concentration. The DNA polymerase is the 'workhorse' in the PCR reaction and the choice of polymerase type can have a significant effect on the success of a reaction. For accurate DNA synthesis, it is essential that the enzyme has 3′–5′ exonuclease (proofreading) activity, which decreases the error rate during PCR amplification (Eckert and Kunkel, 1991; Garcia-Diaz and Bebenek, 2007). *Taq* polymerase isolated from *Thermus aquaticus* is the most commonly used polymerase but it lacks any proofreading activity. In contrast, *Pwo* (*Pyrococcus woesei*) and *Pfu* (*Pyrococcus furiosus*) exhibit good PCR amplification fidelity and proofreading activity (Al-Soud and Rådström, 1998). 'Hot start' polymerases, which were developed to increase the specificity and convenience of the PCR reaction, are another tool worth considering while designing and optimizing an assay (Lebedev *et al.*, 2008). Pre-optimized PCR master mixes are available in which the composition of the reaction is predefined; while each component cannot be optimized in these mixes, they offer advantages in terms of repeatability and preventing variations in reaction concentrations from one reaction to the next.

2.7.4 PCR additives or enhancers

The amplification capacity of PCR can be significantly decreased by inhibitory compounds in the sample (Wiedbrauk *et al.*, 1995; Al-Soud and Rådström, 2001) and by the presence of secondary structures in the DNA template being amplified. Inhibitor compounds (e.g. polyphenols) can often be removed, at least partly, in the sample preparation and extraction steps, but they are not always totally eliminated, and additional purification steps can be time-consuming and have the potential for loss of target nucleic acid (Nolan *et al.*, 2006). These inhibitors can affect the efficiency of the PCR. The use of PCR additives or enhancers such as bovine serum albumin (BSA), dimethyl sulfoxide (DMSO), betaine, glycerol and Tween 20 (see Table 2.1), for example, can help to decrease the effect of inhibitors in the reaction by complexing with them, thereby annulling their effects (e.g. BSA), or by minimizing secondary structures in the DNA template and stabilizing the DNA polymerase (e.g. DMSO, betaine, Tween 20, etc.), thereby positively influencing the reaction fidelity (Demeke and Adams, 1992; Henke *et al.*, 1997; Al-Soud and Rådström, 2001); however, such additives can also have inhibitory effects at high concentration or with some reaction mixes.

Table 2.1. A table of PCR additives showing working concentrations and beneficial effects on amplification.

Additive	Final concentration	Effect
Betaine	0.8–1.3 M	Stabilizes *Taq* polymerase, suppresses secondary structure formation, good for GC-rich templates
BSA (bovine serum albumin)	0.1–0.8 µg/µl	Increases efficiency, 'relieves' inhibition. Can cause inhibition of some reaction mixes
DMSO (dimethyl sulfoxide)	1–5% final	Reduces secondary structure, good for GC-rich templates. Inhibitory at high concentration
Formamide	0.1–2% final	Reduces secondary structure, increases reaction stringency. Inhibitory at high concentration
TMAC (tetramethyl ammonium chloride)	15–100 nM	Increases reaction stringency (very toxic!)
Triton X-100/Tween-20	0.1–1% final	Stabilizes *Taq* polymerase, suppresses secondary structure formation, increases non-specific product formation

Protocols

Contents

Protocol 2.1

One-step PCR to detect potyviruses (Marie-Jeanne *et al.*, 2000)

Materials

- Reagents for conventional RT-PCR: commercial One-Step PCR master mix that contains (or is added separately): reverse transcriptase, *Taq* DNA polymerase, PCR buffer, $MgCl_2$ and dNTPs; and primers and molecular biology grade water. BSA (bovine serum albumin) can be added to reduce the effect of inhibitors found in the nucleic acid extract
- PCR 200 µl reaction tubes
- Pipettors – three separate sets for setting up the reactions, adding RNA to the reactions and adding the RT-PCR reaction to the agarose gel
- Filter tips for pipettors
- RNA sample (can be extracted using a variety of methods, such as phenol/chloroform extraction or commercial-based methods such as a column kit, which uses a silica membrane for purification (e.g. RNeasy, Qiagen).
- Vortex and microcentrifuge
- PCR thermal cycler
- Gel buffer such as Tris-acetate-EDTA (TAE)
- Agarose
- Nucleic acid stain such as SYBR® Safe (ThermoFisher Scientific) or ethidium bromide
- Microwave for melting agarose in the gel buffer
- Molecular marker/ladder and loading dye
- Gel apparatus and electrophoresis unit
- Gel visualization apparatus that emits the correct wavelength (depending on nucleic acid stain used) to detect the DNA (preferably with the capability to capture images of the gel)

Method

1. Calculate the volume of components needed for the number of samples and controls being tested (with samples preferably tested in duplicate). The master mix should cover the number of samples, a no-template control (water), at least two positive controls (see Note 1 below) and, if possible, a healthy plant negative control. Always calculate one more sample than is required to allow for pipetting error.
2. Remove all reagents from freezer and thaw on ice, making sure to keep the RNA samples separate from the rest of the components to avoid contamination. Mix reagents well using a vortex and spin down briefly using a microcentrifuge.
3. Prepare the master mix as shown in Protocol 2.1, Table 1.
4. Fill the required amount of PCR reaction tubes (same as sample amount) with 18 µl of the master mix.
5. Using separate sets of pipettors, add 2 µl of the RNA samples, a no-template control, a negative control and positive controls to the PCR reaction tubes. Add the positive controls last to avoid contaminating the samples.
6. Seal the tubes well, invert them and ensure the RNA is mixed into the master mix. Spin down the tube contents briefly.
7. Place the tubes in the PCR thermal cycler.
8. Run the PCR thermal cycler according to the following conditions: 50°C for 15 min, 94°C for 2 min, followed by 40 cycles of 94°C for 30 s, 50°C for 30 s and 72°C for 30 s. End the PCR reaction with a 7 min extension at 72°C. The samples should then be cooled down to room temperature or below. Samples can be kept in the fridge for later analysis or added directly to the gel.

Protocol 2.1, Table 1. Preparation/composition of one-step PCR master mix.

Reagents	Volume per reaction (µl)	Master mix (µl)[a]
Molecular biology grade water	2.6	3.2
2× Commercial RT-PCR master mix	10.0	20.0
10 mg/ml BSA (optional)	1.0	2.0
5 µM Oligo 1n (Marie-Jeanne *et al.*, 2000)[b]	2.0	4.0
5 µM Oligo 2n (Marie-Jeanne *et al.*, 2000)[b]	2.0	4.0
Reverse transcriptase/polymerase enzyme mix	0.4	0.8
RNA	2.0	–
Total volume	20.0	36.0

[a] In this example, only one sample is calculated.
[b] See Note 2 below.

9. While the samples are in the PCR thermal cycler, a 1.5% agarose gel can be made. The agarose will need to be melted in the gel buffer by boiling in a microwave. Once the agarose buffer has cooled down enough to touch the bottle, the nucleic acid stain can be added to the liquid agarose buffer. Mix the stain gently into the agarose buffer to avoid creating bubbles.
10. The agarose–buffer–stain mix can be poured into a gel apparatus and be left to solidify. Make sure the comb used contains the required amount of wells for each of the samples as well as a well left over for the molecular marker ladder.
11. Once the PCR thermal cycling has finished, the PCR tubes can be removed and taken to the agarose gel loading area.
12. Place the solidified gel into the gel tank and completely submerge in gel buffer.
13. The samples should be mixed to a 1× ratio with loading dye. Add the molecular marker ladder to the well/s to the side/s of the samples. Pipette each sample into the gel wells.
14. Run the gel apparatus tank for 40 min at 90 V using the electrophoresis unit.
15. Once the dye front has at least reached halfway down the gel, the gel can be placed on the gel visualization apparatus. The expected product size should be ~330 bp, which can be compared with the molecular marker.
16. Positive samples can be extracted from the gel and sequenced, or further testing of potyviruses can be undertaken.

Notes

1. It is prudent to include at least two different positive controls when testing with generic primers to give the diagnostician more confidence that the generic primers are indeed able to detect multiple viruses from that group.

2. The two primers used are:

Oligo 1n (5′-ATGGTHTGGTGYATHGARAAYGG-3′) and
Oligo 2n (5′-TGCTGCKGCYTTCATYTG-3′),
where A = adenine, T = thymidine, C = cytosine, G = guanidine, H = A + T + C, K = T + G, R = A + G, Y = C + T.

Protocol 2.2

Nested-PCR to detect phytoplasmas

Materials

- Reagents for conventional PCR: 1 × *Taq* polymerase buffer containing 1.5 mM $MgCl_2$, 0.5 µM of each primer (see Notes 1 and 2 below), 200 µM dNTPs, 1 unit *Taq* DNA polymerase
- DNA template
- All other materials listed in Protocol 2.1

Method

1. Calculate the volume of components needed for the number of samples and controls being tested (with samples preferably tested in duplicate). The master mix should cover the number of samples, a no-template control (water), a positive control and, if possible, a healthy plant negative control. Always calculate one more sample than is required to allow for pipetting error. For the nested PCR, two PCR reactions are required – the second PCR uses the first PCR reaction as the DNA template.
2. Calculate the reagents required as in the master mix table (Table 1) in Protocol 2.1. A 20 µl reaction mixture is composed as follows: 1 × *Taq* polymerase buffer containing 1.5 mM $MgCl_2$, 0.5 µM of each primer, 200 µM dNTPs, 1 unit *Taq* DNA polymerase and 2 µl DNA template.
3. Follow steps 4–7 of Protocol 2.1.
4. Run the PCR thermal cycler according to the following conditions: 5 min at 94°C, 40 cycles of 30 s at 94°C, 30 s at either 53°C (P1/P7 primers – first PCR) or 50°C (R16F2n/R16R2 primers – second PCR), 1 min at 72°C, followed by a final extension for 10 min at 72°C. For nested PCR, the first-stage PCR products may be used

directly as template in the second-stage PCR or can be diluted 1:10 to 1:25 (v/v) in water prior to re-amplification.

5. Follow steps 9–16 of Protocol 2.1. The P1/P7 and R16F2n/R16R2 primers produce a 1800 bp and 1250 bp amplicon, respectively.

Notes

1. The PCR primers used in the first-stage PCR of this assay are P1 (Deng and Hiruki, 1991) and P7 (Schneider *et al.*, 1995), where:

P1 is (5′-AAGAGTTTGATCCTGGCTCAGGATT-3′);
P7 is (5′-CGTCCTTCATCGGCTCTT-3′); and
A = adenine, T = thymidine, C = cytosine, G = guanidine.

2. The PCR primers used in the second-stage PCR of this assay are R16F2n (Gundersen and Lee, 1996) and R16R2 (Lee *et al.*, 1993):

R16F2n is (5′-GAAACGACTGCTAAGACTGG-3′) and has three nucleotides added to the 5′-end of the R16F2 primer of Lee *et al.* (1993);
R16R2 is (5′-TGACGGGCGGTGTGTACAAACCCCG-3′).

References

Al Rwahnih, M., Daubert, S., Golino, D. and Rowhani, A. (2009) Deep sequencing analysis of RNAs from a grapevine showing Syrah decline symptoms reveals a multiple virus infection that includes a novel virus. *Virology* 387, 395–401.

Al-Soud, A. and Rådström, P. (1998) Capacity of nine thermostable DNA polymerases to mediate DNA amplification in the presence of PCR-inhibiting samples. *Applied and Environmental Microbiology* 64, 3784–3753.

Al-Soud, W.A. and Rådström, P. (2001) Purification and characterisation of PCR inhibitory components in blood cells. *Journal of Clinical Microbiology* 39, 485–493.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *Journal of Molecular Biology* 215, 403–410.

Alvarez, A.M. (2004) Integrated approaches for detection of plant pathogenic bacteria and diagnosis of bacterial diseases. *Annual Review of Phytopathology* 42, 339–366.

Andreson, R., Möls, T. and Remm, M. (2008) Predicting failure rate of PCR in large genomes. *Nucleic Acids Research* 36(11), e66.

Avrova, A.O., Hyman, L.J., Toth, R.L. and Toth, I.K. (2002) Application of amplified fragment length polymorphism fingerprinting for taxonomy and identification of the soft rot bacteria *Erwinia carotovora* and *Erwinia chrysanthemi*. *Applied and Environmental Microbiology* 68, 1499–1508.

Boonham, N., González Pérez, L., Mendez, M.S., Peralta, E.L., Blockely, A., Walsh, K., Barker, I. and Mumford, R.A. (2004) Development of a real-time RT-PCR assay for the detection of potato spindle tuber viroid. *Journal of Virological Methods* 116, 139–146.

Boonham, N., Tomlinson, J. and Mumford, R.A. (2007) Microarrays for rapid identification of plant viruses. *Annual Review of Phytopathology* 45, 307–328.

Boyer, S.L., Flechtner, V.R. and Johansen, J.R. (2001) Is the 16S–23S rRNA internal transcribed spacer region a good tool for use in molecular systematics and population genetics? A case study in cyanobacteria. *Molecular Biology and Evolution* 18, 1057–1069.

Chapman, J.R. Taylor, R.K., Weir, B.S., Romberg, M.K., Vanneste, J.L., Luck, J. and Alexander, B.J.R. (2012) Phylogenetic relationships among global populations of *Pseudomonas syringae* pv. *actinidiae*. *Phytopathology* 102, 1034–1044.

Chimento, A., Cacciola, S.O. and Garbelotto, M. (2012) Detection of mRNA by reverse-transcription PCR as an indicator of viability in *Phytophthora ramorum*. *Forest Pathology* 42, 14–21.

Coetzee, B., Freeborough, M.J., Maree, H.J., Celton, J.M., Ress, D.J. and Burger, J.T. (2010) Deep-sequencing analysis of viruses infecting grapevines: viome of a vineyard. *Virology* 400, 157–163.

Collado-Romero, M., Mercado-Blanco, J., Olivares-García, C. and Jiménez-Diaz, R.M. (2008) Phylogenetic analysis of *Verticillium dahliae* vegetative compatibility groups. *Phytopathology* 98, 1019–1028.

Crews, N., Wittwer, C. and Gale, B. (2008) Continuous-flow thermal gradient PCR. *Biomed Microdevices* 10, 187–195.

Cullen, D.W. and Lees, A.K. (2007) Detection of the *nec1* virulence gene and its correlation with pathogenicity in *Streptomyces* species on potato tubers and in soil using conventional and real-time PCR. *Journal of Applied Microbiology* 102, 1082–1084.

Demeke, T. and Adams, R.P. (1992) The effects of plant polysaccharides and buffer additives on PCR. *BioTechniques* 12, 332–334.

Deng, S. and Hiruki, C. (1991) Amplification of 16S rRNA genes from culturable and non-culturable Mollicutes. *Journal of Microbiological Methods* 14, 53–61.

Dixon, L.J., Schlub, R.L., Pernezny, K. and Datnoff, L.E. (2009) Host specialization and phylogenetic diversity of *Corynespora cassiicola*. *Phytopathology* 99, 1015–1027.

Don, R.H., Cox, P.T., Wainwright, B.J., Baker, K. and Mattick, J.S. (1991) Touchdown PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Research* 19(14), 4008.

Eckert, K.A. and Kunkel, T.A. (1991) DNA polymerase fidelity and polymerase chain reaction. *PCR Methods and Applications* 1, 17–24.

Einax, E. and Voigt, K. (2003) Oligonucleotide primers for the universal amplification of β-tubulin genes facilitate phylogenetic analyses in the regnum Fungi. *Origins, Diversity and Evolution* 3, 185–194.

Foster, S.J., Ashby, A.M. and Fitt, B.D.L. (2002) Improved PCR-based assays for presymptomatic diagnosis of light leaf spot and determination of mating type of *Pyrenopeziza brassicae* on winter oilseed rape. *European Journal of Plant Pathology* 108, 379–383.

Fraaije, B.A., Lovell, D.J., Rohel, E.A. and Hollomon, D.W. (1999) Rapid detection and diagnosis of *Septoria tritici* epidemics in wheat using a polymerase chain reaction PicoGreen assay. *Journal of Applied Microbiology* 86, 701–708.

Garcia-Diaz, M. and Bebenek, K. (2007) Multiple functions of DNA polymerases. *Critical Reviews in Plant Sciences* 26, 105–122.

Geiser, D.M., Jiménez-Gasco, M., Kang, S., Makalowska, I., Veeraraghavan, N., Ward, T.J., Zhang, N., Kuldau, G.A. and O'Donnell, K. (2004) Fusarium-ID v.1.0: a DNA sequence database for identifying *Fusarium*. *European Journal of Plant Pathology* 110, 473–479.

Glienke, C., Pereira, O.L., Stringari, D., Fabris, J., Kava-Cordeiro, V., Galli-Terasawa, L., Cunnington, J., Shivas, R.G., Groenewald, J.Z. and Crous, P.W. (2011) Endophytic and pathogenic *Phyllosticta* species, with reference to those associated with citrus black spot. *Persoonia* 26, 47–56.

Gundersen, D.E. and Lee, I.-M. (1996) Ultrasensitive detection of phytoplasmas by nested-PCR assays using two universal primer pairs. *Phytopathologia Mediterranea* 35, 144–151.

Guo, Y.H., Cheng, Z.M. and Walla, J.A. (2003) Rapid PCR based detection of phytoplasmas from infected plants. *HortScience* 38, 1134–1136.

Haible D., Kober, S. and Jeske, H. (2006) Rolling circle amplification revolutionizes diagnosis and genomes of geminiviruses. *Journal of Virological Methods* 135, 9–16.

Harper, S.J., Ward, L.I. and Clover, G.R.G. (2010) Development of a LAMP and real-time PCR methods for the rapid detection of *Xylella fastidiosa* for quarantine and field applications. *Phytopathology* 100, 1282–1288.

Harper, S.J., Delmilglio, C., Ward, L.I. and Clover, G.R.G. (2011) Detection of *Tomato blackring virus* by real-time one-step RT-PCR. *Journal of Virological Methods* 171, 190–194.

Hecker, K.H. and Roux, K.H. (1996) High and low annealing temperatures increase both specificity and yield in touchdown and stepdown PCR. *BioTechniques* 20, 478–485.

Henke, W., Herdel, K., Jung, K., Schnorr, D. and Loening, S.A. (1997) Betaine improves the PCR amplification of GC-rich DNA sequences. *Nucleic Acids Research* 25, 3957–3958.

Hirsch, P.R., Mauchline, T.H., Mendum, T.A. and Kerry, B.R. (2000) Detection of the nematophagous fungus *Verticillium chlamydosporium* in nematode-infested plant roots using PCR. *Mycological Research* 104, 435–439.

Hodgetts, J., Boonham, N., Mumford, R., Harrison, N. and Dickinson, M. (2008) Phytoplasma phylogenetics based on analysis of *secA* and 23S rRNA gene sequences for improved resolution of candidate species of '*Candidatus* Phytoplasma'. *International Journal of Systematic and Evolutionary Microbiology* 58, 1826–1837.

IRPCM Phytoplasma/Spiroplasma Working Team – Phytoplasma Taxonomy Group (2004) '*Candidatus* Phytoplasma', a taxon for the wall-less, non-helical prokaryotes that colonize plant phloem and insects. *International Journal of Systematic and Evolutionary Microbiology* 54, 1243–1255.

Ishii, K. and Fukui, M. (2001) Optimization of annealing temperature to reduce bias caused by a primer mismatch in multitemplate PCR. *Applied and Environmental Microbiology* 67, 3753–3755.

Kato, S., Shimamoto, Y. and Mikami, T. (1995) The apple mitochondrial atp9 gene: RNA editing and co-transcription with exons a and b of the *nad5* gene. *Physiologia Plantarum* 93, 572–575.

Kunz, B.A. and Kohalmi, S.E. (1991) Modulation of mutagenesis by deoxyribonucleotide levels. *Annual Review of Genetics* 25, 339–359.

Lebedev, A.V., Paul, N., Yee, J., Timoshchuk, V.A., Shum, J., Miyagi, K., Kellum, J., Hogrefe, R.I. and Zon, G. (2008) Hot start PCR with heat-activated primers: a novel approach for improved PCR performance. *Nucleic Acids Research* 36(20), e131.

Lee, I.-M., Hammond, R.W., Davis, R.E. and Gundersen, D.E. (1993) Universal amplification and analysis of pathogen 16S rDNA for classification and identification of mycoplasma-like organisms. *Phytopathology* 83, 834–842.

Lee, I.-M., Gundersen-Rindal, D.E., Davis, R.E. and Bartoszyk, I.M. (1998) Revised classification scheme of phytoplasmas based on RFLP analyses of 16S rRNA and ribosomal protein gene sequences. *International Journal of Systematic Bacteriology* 48, 1153–1169.

Lee, I.-M., Bottner-Parker, K.D., Zhao, Y., Davis, R.E. and Harrison, N.A. (2010) Phylogenetic analysis and delineation of phytoplasmas based on *secY* gene sequences. *International Journal of Systematic and Evolutionary Microbiology* 60, 2887–2897.

Li, X. and De Boer, S.H. (1995) Selection of polymerase chain reaction primers from an RNA intergenic spacer region for specific detection of *Clavibacter michiganensis* subsp. *sepedonicus*. *Phytopathology* 85, 837–842.

Lievens, B., Van Baarlen P., Verreth C., Van Kerckhove, S., Rep, M. and Thomma, B.P.H.J. (2009) Evolutionary relationships between *Fusarium oxysporum* f. sp *lycopersici* and *F. oxysporum* f. sp *radicis-lycopersici* isolates inferred from mating type, elongation factor-1 alpha and exopolygalacturonase sequences. *Mycological Research* 113, 1181–1191.

Lim, P.-O. and Sears, B.B. (1992) Evolutionary relationships of a plant-pathogenic mycoplasmalike organism and *Acholeplasma laidlawii* deduced from two ribosomal protein gene sequences. *Journal of Bacteriology* 174, 2606–2611.

López, M.M., Bertolini, E., Olmos, A., Caruso, P., Gorris, M.T., Llop, P., Penyalver, R. and Cambra, M. (2003) Innovative tools for detection of plant pathogenic viruses and bacteria. *International Microbiology* 6, 233–243.

Maes, M., Garbeva, P. and Crepel, C. (1996) Identification and sensitive endophytic detection of the fire blight pathogen *Erwinia amylovora* with 23S ribosomal DNA sequences and the polymerase chain reaction. *Plant Pathology* 45, 1139–1149.

Makarova, O., Contaldo, N., Paltrinieri, S., Kawube, G., Bertaccini, A. and Nicolaisen, M. (2012) DNA barcoding for identification of '*Candidatus* Phytoplasmas' using a fragment of the elongation factor Tu gene. *PLoS ONE* 7(12), e52092.

Manceau, C. and Horvais, A. (1997) Assessment of genetic diversity among strains of *Pseudomonas syringae* by PCR-restriction fragment length polymorphism analysis of rRNA operons with special emphasis on *Pseudomonas syringae* pv. tomato. *Applied and Environmental Microbiology* 63, 498–505.

Marie-Jeanne, V., Ioos, J., Peyre, J., Alliot, B. and Signoret, P. (2000) Differentiation of Poaceae potyviruses by reverse transcription-polymerase chain reaction and restriction analysis. *Phytopathology* 148, 141–151.

Markoulatos, P., Siafakas, N. and Moncany, M. (2002) Multiplex polymerase chain reaction: a practical approach. *Journal of Clinical Laboratory Analysis* 16, 47–51.

Martin, F.N., Coffey, M.D., Zeller, K., Hamelin, R.C., Tooley, P., Garbelotto, M., Hughes, K.J., Kubisiak, T., Bilodeau, G.J., Levy, L. *et al*. (2009) Evaluation of molecular markers for *Phytophthora ramorum* detection and identification: testing for specificity using a standardized library of isolates. *Phytopathology* 99, 390–403.

Martini, M., Lee, I.-M., Bottner, K.D., Zhao, Y., Botti, S., Bertaccini, A., Harrison, N.A., Carraro, L., Marcone, C., Khan, A.J. and Osler, R. (2007) Ribosomal protein gene-based phylogeny for

finer differentiation and classification of phytoplasmas. *International Journal of Systematic and Evolutionary Microbiology* 57, 2037–2051.

Măruţescu, L., Manole, F. and Sesan, T. (2009) Genetic characterization of *Erwinia amylovora* strains by random amplified polymorphic DNA fragments. *Romanian Archives of Microbiology and Immunology* 68, 166–170.

Menzel, W., Jelkmann, W. and Maiss, E. (2002) Detection of four apple viruses by multiplex RT-PCR assays with coamplification of plant mRNA as internal control. *Journal of Virological Methods* 99, 81–92.

Moleleki, L.N., Onkendi, E.M., Mongae, A. and Kubheka, G.C. (2013) Characterisation of *Pectobacterium wasabiae* causing blackleg and soft rot diseases in South Africa. *European Journal of Plant Pathology* 135, 279–288.

Mulè, G., Susca, A., Stea, G. and Moretti, A. (2004) Specific detection of the toxigenic species *Fusarium proliferatum* and *F. oxysporum* from asparagus plants using primers based on calmodulin gene sequences. *FEMS Microbiology Letters* 230, 235–240.

Nolan, T., Hands, R.E., Ogunkolade, W. and Bustin, S.A. (2006) SPUD: a quantitative PCR assay for the detection of inhibitors in nucleic acid preparations. *Analytical Biochemistry* 351, 308–310.

Oguiza, J.A., Rico, A., Rivas, L.A., Sutra, L., Vivian, A. and Murillo, J. (2004) *Pseudomonas syringae* pv. *phaseolicola* can be separated into two genetic lineages distinguished by the possession of the phaseolotoxin biosynthetic cluster. *Microbiology* 150, 473–482.

Palacio-Bielsa, A., Cambra, M.A. and López, M.M. (2009) PCR detection and identification of plant pathogenic bacteria: updated review of protocols (1989–2007). *Journal for Plant Pathology* 91, 249–297.

Pitman, A.R., Harrow, S.A. and Visnovsky, S.B. (2010) Genetic characterisation of *Pectobacterium wasabiae* causing soft rot disease of potato in New Zealand. *European Journal of Plant Pathology* 126, 423–435.

Puchta, H. and Sanger, H.L. (1989) Sequence analysis of minute amounts of viroid RNA using the polymerase chain reaction (PCR). *Archives of Virology* 106, 335–340.

Puławska, J. and Sobiczewski, P. (2012) Phenotypic and genetic diversity of *Erwinia amylovora* the causal agent of fire blight. *Trees* 26, 3–12.

Roux, K.H. (1995) Optimization and troubleshooting in PCR. *Genome Research* 4, S185–S194.

Sarkar, S.F. and Guttman, D.S. (2004) Evolution of the core genome of *Pseudomonas syringae*, a highly clonal, endemic plant pathogen. *Applied and Environmental Microbiology* 70, 1999–2012.

Schneider, B., Seemüller, E., Smart, C.D. and Kirkpatrick, B.C. (1995) Phylogenetic classification of plant pathogenic mycoplasma like organisms or phytoplasmas. In: Razin, S. and Tully, J.G. (eds) *Molecular and Diagnostic Procedures in Mycoplasmology, Vol. 1*. Academic Press, San Diego, California, pp. 369–380.

Schoch, C.L., Seifert, K.A., Huhndorf, S., Robert, V., Spouge, J.L., Levesque, C.A., Chen, W. and Fungal Barcoding Consortium (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences of the United States of America* 109, 6241–6246.

Scortichini, M., Dettori, M.T., Marchesi, U., Palombi, M.A. and Rossi, M.P. (1998) Differentiation of *Pseudomonas avellanae* strains from Greece and Italy by rep-PCR genomic fingerprinting. *Journal of Phytopathology* 146, 417–420

Shampo, M.A. and Kyle, R.A. (2002) Kary B. Mullis – Nobel Laureate for procedure to replicate DNA. *Mayo Clinic Proceedings* 77, 606.

Sheridan, G.E.C., Masters, C.I., Shallcross, J.A. and Mackey, B.M. (1998) Detection of mRNA by reverse transcription-PCR as an indicator of viability in *Escherichia coli* cells. *Applied and Environmental Microbiology* 64, 1313–1318.

Stemmer, W.P., Crameri, A., Ha, K.D., Brennan, T.M. and Heyneker, H.L. (1995) Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* 164, 49–53.

Taylor, R.K., Chapman, J.R., Romberg, M.K., Weir, B.S., Vanneste, J.L., Everett, K.R., Ward, L.I., Liefting, L.W., Lebas, B.S.M. and Alexander, J.R. (2014) Diagnostic challenges for the detection of emerging pathogens: a case study involving incursion of *Pseudomonas syringae* pv. *actinidiae* in New Zealand. In: Lodovica Gullino, M. and Bonants, P.J.M. (eds) *Detection and Diagnostics of Plant Pathogens. Volume 5, Plant Pathology in the 21st Century*. Springer, Dodrecht, the Netherlands, pp. 71–85.

Tomlinson, J.A., Barker, I. and Boonham, N. (2007) Faster, simpler, more-specific methods for improved molecular detection of *Phytopthora ramorum* in the field. *Applied and Environmental Microbiology* 73, 4040–4047.

Validov, S.Z., Kamilova, F.D. and Lugtenberg, B.J.J. (2011) Monitoring of pathogenic and non-pathogenic *Fusarium oxysporum* strains during tomato plant infection. *Microbial Biotechnology* 4, 82–88.

Wanner, L.A. (2004) Field isolates of *Streptomyces* differ in pathogenicity and virulence on radish. *Plant Disease* 88, 785–796.

Wanner, L.A. (2006) A survey of genetic variation in *Streptomyces* isolates causing potato common scab in the United States. *Phytopathology* 96, 1363–1371.

Ward, L.I., Burnip, G.M., Liefting, L.W., Harper, S.J. and Clover, G.R.G. (2011) First report of *Grapevine yellow speckle viroid 1* and *Hop stunt viroid* in grapevine (*Vitis vinifera*) in New Zealand. *Plant Disease* 95, 617.

Warrens, A.N., Jones, M.D. and Lechler, R.I. (1997) Splicing by overlap extension by PCR using asymmetric amplification: an improved technique for the generation of hybrid proteins of immunological interest. *Gene* 186, 29–35.

Watson, J.D. and Crick, F.H. (1953) Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature* 171, 737–738.

Weller, S.A., Elphinstone, J.G., Smith, N.C., Boonham, N. and Stead, D.E. (2000) Detection of *Ralstonia solanacearum* strains with a quantitative, multiplex, real-time, fluorogenic PCR (TaqMan) assay. *Applied and Environmental Microbiology* 66, 2853–2858.

Wiedbrauk, D.L., Werner, J.C. and Drevon, A.M. (1995) Inhibition of PCR by aqueous and vitreous fluids. *Journal of Clinical Microbiology* 33, 2643–2646.

Wittwer, C.T. and Garling, D.J. (1991) Rapid cycle DNA amplification: time and temperature optimization. *Biotechniques* 10, 76–83.

3 Real-time PCR

Maja Ravnikar, Nataša Mehle, Kristina Gruden and Tanja Dreo*

National Institute of Biology, Ljubljana, Slovenia

3.1 Introduction

Real-time PCR has become the PCR method of choice for diagnostic applications for a number of reasons. It is generally more sensitive than conventional PCR and can be better controlled, and these are important factors for diagnostic applications. Real-time PCR is also more robust than conventional PCR owing to a number of factors, including the amplification of small fragment sizes, the implementation of fluorescent detection, which enables the detection of smaller amounts of amplified products, and improved tolerance to inhibiting compounds that are co-purified with the DNA when compared with end-point PCR methods. Further, being a closed tube system helps to prevent post-PCR contamination and resulting false positive results. Finally, because gel electrophoresis is not required, the method can be easily automated, making it suitable for fast, high-throughput testing applications.

3.2 Coupling PCR amplification with the generation of fluorescence

Real-time PCR is a variant of PCR in which the amplification of a product in each cycle is coupled to the generation of fluorescence. Optical systems are built into the thermal cycler and these excite and measure fluorescence in each reaction over time. There are different approaches to translating DNA amplification into fluorescence; the two most commonly used are hydrolysis probes and intercalating dyes, and these are discussed further in this chapter.

*Corresponding author: tanja.dreo@nib.si

One output of real-time PCR is the 'amplification plot' (Fig. 3.1), a graphical output of fluorescence against the number of cycles. The amplification plot of normalized fluorescence in a positive sample is typically of a sigmoidal shape and is characterized by a series of phases: (i) the baseline – the first cycles during which fluorescence is low and is part of the background noise; (ii) an exponential increase corresponding to the exponential amplification of the product and the associated fluorescence, which is followed by a linear increase; and (iii) a plateau phase. The reactions are characterized by the point of the cycling protocol at which the amplification of a target achieves a fixed level of fluorescence, rather than the end point amount of the PCR product. This fluorescence level is defined by a threshold above the background fluorescence. The PCR cycle during which the signal rises above the threshold is named a 'quantification cycle', abbreviated as Cq, which has been proposed as a generic term to be used instead of the proprietary designations of Cp (crossing point), TOF (take-off point), etc. (Bustin *et al.*, 2009). The Cq correlates with the starting template copy number. The higher the concentration of this at the beginning, the faster enough fluorescence is generated to reach the threshold. Consequently, samples with higher target concentrations will have a lower Cq than those with lower target concentrations. In negative samples, there is no amplification, the fluorescence does not increase and no Cq value is obtained.

3.3 Probe-based real-time PCR

The use of probes is a very common way of translating DNA amplification into fluorescence in real-time PCR. The most commonly used probes are hydrolysis TaqMan® probes. These are short oligonucleotides (typically from 18 to 30 bp) complementary to

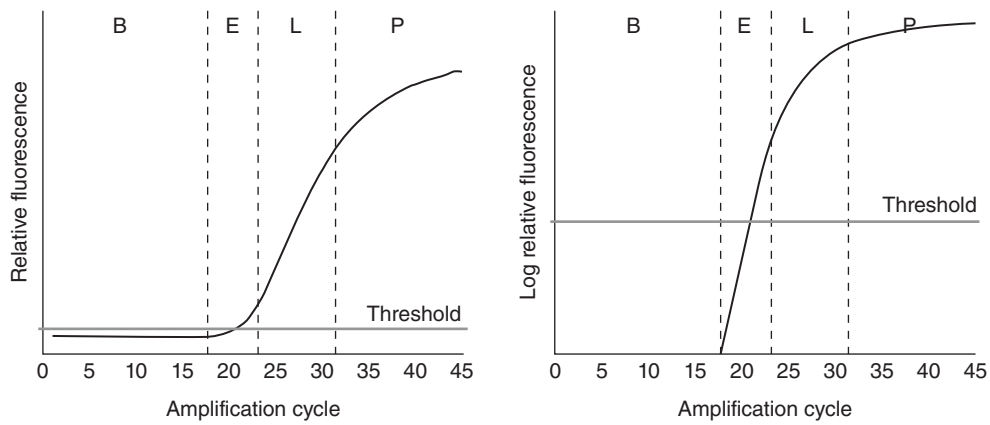


Fig. 3.1. Typical amplification plot of a positive sample showing relative fluorescence over amplification cycles in linear (left) and logarithmic views (right). The amplification can be divided into several phases: the baseline (B), during which the signal is still below the threshold; exponential (E) and linear amplification (L); and a plateau phase (P). The Cq (quantification cycle) of a sample in which the fluorescence of the sample increased above the selected threshold is 20.9 in this case. In real-time PCR, the Cq value correlates with the starting amount of the target. In negative samples, no increase of fluorescence is observed.

the DNA sequence between the two primer annealing sites. TaqMan® probes are labelled with a reporter dye on the 5′ end and a quencher molecule on the 3′ end. The dye and the quencher are chosen in such a way that Förster resonance energy transfer (FRET) occurs between them. FRET has been extensively used in studying protein–protein and DNA–protein interactions and protein conformational changes. When the reporter and the quencher are close, e.g. when the probe is intact, the reporter, initially in its electronic excited state, transfers energy to the quencher through non-radiative dipole–dipole coupling. During extension of the DNA strand by *Taq* polymerase, the 5′→3′ exonuclease activity of the enzyme leads to cleavage of the probe and release of the cut pieces into the solution. In the solution, the distance between the reporter and the quencher increases, so FRET cannot occur, i.e. the energy is no longer transferred from the reporter to the quencher dye. Consequently, the fluorescence of the released reporter dye is of a different wavelength and can be measured separately from that of the reporter dye that is still a part of an intact probe. As with primers, the probe is added to the reaction in excess and new intact probes anneal to the target in each amplification cycle, are cleaved and the released reporter dye contributes to the total measured fluorescence.

For FRET to take place, a compatible acceptor for a certain donor is a molecule whose absorbance spectrum overlaps the emission spectrum of the donor molecule. One of the most frequently used combinations of dyes is 6-carboxyfluorescein (FAM), which acts as the reporter fluorophore, and tetramethylrhodamine (TAMRA), which acts as a quencher molecule. TAMRA is a fluorophore and can also be followed during real-time PCR. In a positive sample, the measured fluorescence of released FAM increases, while the measured fluorescence of TAMRA in the intact probe decreases (Fig. 3.2). A number of other reporter and quencher molecules are now available, and the choice largely depends on the properties of the optical system and the software used, which come calibrated for a certain range of fluorophores. Widely available reporter dyes include, for example, VIC®, JOE (a dimethoxyfluorescein derivative), HEX (hexachlorfluorescein), Yakima Yellow and others. TAMRA also can be used as a reporter dye provided that it is not used as a quencher dye in the same reaction. For several years now, non-fluorescent quenchers have been available and are commonly used, e.g. Black Hole Quenchers (BHQs).

Another hydrolysis probe-based chemistry that is relatively common is that using the TaqMan® MGB (minor groove binder) probes. MGB probes use non-fluorescent quencher dyes and an additional molecule, a minor groove binder, which stabilizes probe duplexes with single-stranded DNA (ssDNA) targets (Kutyavin *et al*., 2000). Because of this stabilizing effect, the probes themselves can be shorter while still having a melting temperature high enough to hybridize efficiently to the template. This enables the design of a useful probe even when the target sequence available for probe design is very short. In addition, MGB probes are more specific for discriminating single base mismatches (Kutyavin *et al*., 2000).

Other types of hydrolysis probes have been developed that usually aim to allow for single nucleotide polymorphism (SNP) identification, i.e. increased discrimination capabilities. Variants more commonly used include TaqMan® locked nucleic acid (LNA) probes, high-affinity RNA analogues with increased mismatch detection, which are particularly suitable for targeting AT (adenine/thymine)-rich target sequences (Koshkin *et al*., 1998),

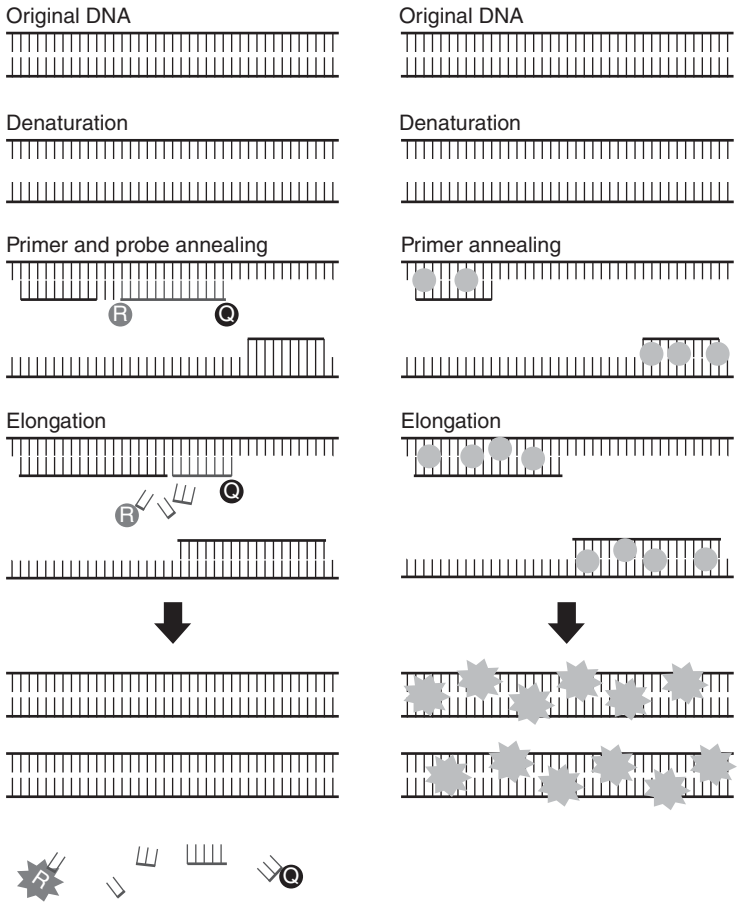


Fig. 3.2. The coupling of PCR amplification with the generation of fluorescence using probes (left) or intercalating dyes (right). The most commonly used probes are hydrolysis (TaqMan®) probes, which are labelled with a reporter dye (left-hand sequence of reactions), and are cleaved during PCR amplification. This releases the reporter dye (R) into the solution, thus increasing its distance from the quencher (Q), so that the reporter dye is no longer quenched. During amplification, the fluorescence increases in each cycle as new probes are cut. The intercalating dyes bind to all of the double-stranded (ds) DNA produced and in such a state emit fluorescence upon excitation. See text for further details.

and molecular beacons (oligonucleotide hybridization probes), in which the signal depends on their conformational change upon hybridization of the probe to the PCR template (Tyagi and Kramer, 1996, 2012). Examples of the use of these probes in the field of plant pathology include phytoplasma detection (Palmano *et al.*, 2015) and the detection and differentiation of *Potato virus Y* (Szemes *et al.*, 2002), respectively.

In general, the use of hybridization probes is preferable over non-probe-based chemistries for the design of highly specific PCR assays. Probes can distinguish between different reaction products obtained using the same primer set or allow for multiplex PCR, in which different PCR products are detected using different dyes (Mackay *et al.*, 2002).

Many other approaches to real-time PCR and fluorescent probes have been described in the literature that follow similar principles to those outlined above for the most commonly used tools, but are used to a lesser extent (for reviews see Buh Gašparič *et al*., 2008; Juskowiak, 2011).

3.4 Non-probe chemistries

DNA-binding dyes change conformation as they bind with double-stranded DNA (dsDNA), which increases their fluorescence and results in signals that are proportional to the amount of DNA at the point of measurement. They are not sequence specific, i.e. they bind to any dsDNA (Fig. 3.2). In the absence of a probe, the specificity of the assay is defined by the primers, reaction mixture and cycling conditions, as in conventional PCR. All dsDNAs, including primer dimers and non-specific products, are detected and contribute to the total fluorescence and thus to the Cq value. The most commonly used intercalating dyes include SYBR® Green and EvaGreen®, but others are available. The identity, or specificity, of the product is confirmed by melting temperature analysis, an additional step in the cycling programme in which the temperature is gradually changed in small increments, during which overall fluorescence changes depend on the amount of dsDNA. The temperature at which the biggest change in fluorescence is observed is defined as the melting temperature (T_m), at which 50% of the product is in the form of dsDNA and 50% in the form of ssDNA. Ideally, the product should yield a sharp peak in the first derivative plot of fluorescence against temperature. The exact T_m depends on the length of the product, its secondary structure, the sequence itself and, with some dyes, the concentration of the target. Depending on the assay design and the dye employed, the resolution can be high, and so allow the discrimination of viral strains through high-resolution melting curve analysis (see, for example, Bester *et al*., 2012; Gori *et al*., 2012).

3.5 Reaction components and cycling programmes

As in conventional PCR, the reaction mixture for real-time PCR contains all the components necessary for PCR amplification: DNA polymerase, deoxynucleotide triphosphates (dNTPs), reaction buffer, primers and the DNA template (sample). In addition, probes are added or, when using intercalating dyes, a suitable dsDNA-binding dye is a part of the reaction.

Pre-prepared master mixes are available that contain DNA polymerase, dNTPs and reaction buffer, and these may improve reproducibility. DNA polymerases are a critical component and can differ in their optimal temperature (commonly 72 or 60°C), processivity, extension rate and accuracy. The use of a thermostable, hot-start polymerase is recommended, but care should be taken to select suitable activation conditions, typically a 2–10 min incubation step at high temperature (95°C) before PCR cycling. Some polymerases have additional activities, e.g. reverse transcription activity and strand displacement activity, which can improve the real-time PCR. Real-time PCR protocols employing faster temperature ramping (e.g. protocols run on SmartCycler® (Cepheid) at maximum temperature ramping) also often require dedicated reaction mixtures and polymerases. As another critical component, $MgCl_2$ ions can be supplied either as a part of pre-prepared master mixes or separately. Most commercial buffers also contain

co-solvents, which can influence template melting properties, PCR enhancers, which increase product yield, and compounds that coat the sides of the PCR tubes so that reagents are not lost through adsorption to the tube walls (Bustin, 2012). Pre-prepared master mixes with an increased resistance to common inhibitors are available (Trombley Hall *et al*., 2013).

A pre-prepared master mix comes with a recommended reaction composition and a cycling protocol suitable for typical applications, although this may have to be adapted depending on the primer and probe melting temperature, and the size of the PCR product. Typically, a laboratory aims to run all real-time PCR at the same cycling, thereby allowing the running of different tests at the same time. Using, for example, the TaqMan® Universal Master Mix II for a typical real-time PCR, the following cycling protocol is commonly applied: 10 min at 95°C (initial template denaturation and activation of the hot-start polymerase), followed by 45 cycles of 15 s at 95°C (denaturation), and 1 min at 60°C (primer annealing and extension). Generally, a two-step cycling protocol which combines the primer annealing and extension in one step is possible if melting temperatures of the primers are all above 60°C; if this is not the case, the annealing and extension steps may have to be separated. When detecting amplification by intercalating dyes, an additional step is added at the end of amplification. This step consists of slow temperature changes, typically between 60 and 90°C, and simultaneous fluorescence measurements that enable the determination of the melting temperature of the product.

Reaction volume in real-time PCR is partially determined by the instrument, with a typical total reaction volume in the range of 10–25 µl. The typical final concentrations of primers and probes are 900 nM for each primer and 200–250 nM for the probe. The larger reaction volumes may allow a larger volume of sample to be analysed. Theoretically, this should improve analytical sensitivity; however, as a larger volume also means more background fluorescence, the expected improved sensitivity is generally not observed in practice.

When targeting RNA organisms, RNA first needs to be (reverse) transcribed into DNA before it can be amplified by PCR. The reverse transcription (RT) reaction can be done in a separate reaction before real-time PCR (two-step RT real-time PCR) or in the same tube as the PCR, in which case the reverse transcription is incubated before the PCR starts (one-step RT real-time PCR). In both cases, specialized pre-prepared master mixes are available.

It is worth noting that the fluorescence emission of any molecule, including reporter dyes, is dependent on environmental factors such as the pH of a solution and the salt concentration. Consequently, using reaction mixtures with different compositions can influence the fluorescence, lead to different baseline fluorescence (noise) and, consequently, influence the signal to noise ratio. Any change in reaction mixture is advised to be checked experimentally for performance.

3.6 Real-time PCR instruments

Real-time PCR instruments consist of: (i) a thermal system to perform temperature cycling; (ii) an optical system with a light source to excite light from the fluorophores (laser or LED); and (iii) a system to detect and measure the generated fluorescence

(photodiodes, photomultiplier tubes). Instruments differ in their technical specifications; for example, the approach to recording data (reading several or all wells at the same time), spectral range (defining which and how many dyes can be used) and temperature uniformity. The temperature ramping speed, which is crucial for faster reactions, can differ widely from ramping at the level of classical PCR cyclers (e.g. +0.8°C/s and –1.6°C/s in 9600 Emulation Mode on a 7900HT Fast Real Time PCR System from Life Technologies) to rapid changes of up to 20°C/s and more (e.g. LightCycler 2.0 from Roche and Rotor-Gene Q from Qiagen), with most commonly used real-time PCR instruments having an average temperature ramping of 3°C/s. Among the reaction formats, 96 well and 384 well plates are most common, but other formats are available (e.g. single tubes and 8 well strips, 32, 48 and 72 well plates and 1536 array based). Some instruments have been designed to be portable, e.g. SmartCycler® (Cepheid).

Proprietary software is used to control the instrument operation and to collect and analyse the data generated. The software can differ in normalization and quantification approaches. In addition to proprietary software linked to a specific thermal cycler, instrument-independent programs are available for various operations, from baseline and threshold settings to more advanced calculations of amplification efficiency and quantification. A review of different curve analysis methods for gene expression studies has recently been published by Ruijter *et al.* (2013); this compares the performance of various publicly available curve analysis methods using a previously published large clinical data set.

The transfer of assays among different instruments is possible and often straightforward, although critical parameters may need to be adapted and verified. Protocols developed under slower amplification conditions can often be transferred to faster instruments; however, optimization is required (Bentley *et al.*, 2005) to exploit fully the possibility of faster temperature ramping. For example, the transfer of the Ams real-time PCR assay for the detection of *Erwinia amylovora* from the 7900 HT SDS (Life Technologies) to the SmartCycler® (Cepheid) instrument was successful, decreasing the time needed for amplification from 1 h 40 min to 30 min, with no impairment of analytical sensitivity (Dreo *et al.*, 2012). This transfer included adaptation of the amplification protocol and optimization of: (i) primers and probe concentrations; (ii) annealing temperature and time; (iii) $MgCl_2$ concentration; (iv) temperature ramping rate; and (v) sample volume (see also Protocol 3.1).

3.7 Selecting and designing an assay

Many assays have been described and evaluated for the detection of plant pathogens. In choosing a suitable assay, what is needed from the assay is taken into account (i.e. the assay should be fit for purpose) and as much information as possible is collected from original published data on the assay, together with separately deposited validation data if this exists (e.g. data in the European and Mediterranean Plant Protection Organization (EPPO) database on diagnostic expertise: http://dc.eppo.int/validationlist.php). *In silico* analysis of the secondary structures of the primers and probe and of their specificity through BLAST (the Basic Local Alignment Search Tool from GenBank, the

US National Institutes of Health genetic sequence database available from NCBI, the US National Center for Biotechnology Information) analysis may save a lot of experimental effort. If there is no suitable real-time PCR assay yet available, it may be necessary to design an assay.

3.7.1 Choosing a target sequence

To be able to choose a target sequence and develop an assay fit for purpose, the first step is to define the scope of the assay accurately: (i) which organism is to be detected; (ii) in which matrix (sample) it is to be detected; and (iii) any specific requirement related to the circumstances of the use of the test (EPPO, 2010). For the detection of plant pathogens, particularly those with quarantine status, the aim is usually low-level detection and the target is a taxonomically defined unit or its subgroup, e.g. a species or a pathovar. Sometimes, the interest is in a subset of the target isolates with particular characteristics. These characteristics can have a background in a known DNA sequence of a limited size, e.g. the Ti and Ri plasmids that confer virulence in *Agrobacterium*. In many cases, however, the causal relationship between the observed characteristic of the harmful organism and its DNA is complex or not yet described.

Choosing the target sequence is the first and the most important step in the assay design. The aim is to try to find a sequence that is specific for all the isolates of the target organism, thus ensuring that all isolates will be detected without the sequence sharing significant similarities with the host plant DNA, related organisms or any other organisms present in the chosen matrix. The more information is available on the population structure and phylogeny of the target organism and of host plant, along with its microflora, the better.

Historically, many PCR assays, for example, in bacteriology, targeted amplified genomic regions such as the 16S rDNA and ribosomal internal transcribed spacer (ITS) regions, and it is not surprising that many real-time PCR assays also target these regions. Having multiple copies of the target increases the reliability of low-level detection because it substitutes the uncertainty of having to identify a single copy of a gene or single genome with the consistency and reliability of being able to target tens of copies specified within that single genome. Sequences found on plasmids can be useful as they are also often present in multiple copies, but care should be taken in interpretation, keeping in mind that plasmids may be lost from the genome, especially when the target organism is under unfavourable conditions. Depending on the target organism, the discriminatory value of 16S rDNA can be poor, and assays are being developed that target other sequences, e.g. housekeeping genes used in phylogenetic research and virulence genes. If a test is also to be used for quantification, a variable number of target copies in isolates can lead to inaccurate calculations, so single copy genes are preferred for some applications. With the increase in available genome sequencing data, a more informed approach to the selection of potentially new target sequences is possible (e.g. Gabriel *et al.*, 2006; Pritchard *et al.*, 2013).

BLAST analysis of DNA sequences publicly available in databases (such as NCBI's GenBank, the European Nucleotide Archive, the DNA Data Bank of Japan) and

multiple sequence alignment are powerful tools that allow choice of the possible template sequence of a size suitable for designing an assay. In addition to BLAST analysis of primers, the probe and the whole PCR product, many other freely available tools can be used for a more thorough *in silico* analysis of both the target sequence and the designed oligonucleotides. In general, sequences that may create strong secondary structures, including hairpins, self-dimers and heterodimers, are best avoided.

3.7.2 Designing primers and probe

A selected sequence can be used as input for assay design software that also takes into account the technical requirements for primers and probes. A hydrolysis TaqMan® probe is designed to be as short as possible (less than 30 nucleotides). Most importantly, its T_m should be 4–10°C above the T_m values of the two primers to ensure that it anneals to the template during amplification and can so can be cleaved by the polymerase. It is recommended that both primer and probe have GC (guanine/cytosine) content between 30 and 80%, and runs of the same nucleotide should be avoided to minimize the possibility of secondary structure formation. Primers are usually designed with the target melting temperature between 58 and 60°C, and less than 2°C difference between the two primers, and of length 15–30 nucleotides. It is recommended that primers do not contain more than 2 Gs or Cs in the last five bases at the 3′ end to ensure their stability and thus the specificity of their annealing to the template. The probe should not contain G at the 5′ end because G itself can act as a quencher to some reporter dyes, and thereby cause a reduction in the fluorescent signal. For a high-throughput laboratory environment, it is useful to design assays with the same running conditions, which allows for parallel testing of multiple assays on the same reaction plate.

3.8 Sample preparation and controls

Sampling and sample preparation are often the biggest source of uncertainty in the determination of pathogen presence because they remain largely non-standardized even for latent infection testing. Cross-contamination of samples can occur during sampling and sample preparation in the laboratory, as well as during the further steps of nucleic acid extraction and real-time PCR set-up.

Typical steps in the preparation of a sample include subsampling and preparation of the plant extract by macerating or commuting plant tissue in a buffer solution, after which process a defined amount of a plant extract is used for nucleic acid extraction.

The use of controls enables the identification of problems and following of reaction efficiency; it includes both positive and negative controls. In the absence of reference materials, positive controls are prepared and characterized in-house. Controls are prepared and analysed in such a way as to provide information on the performance of the nucleic acid extraction and real-time PCR step separately. All controls are processed in the same way and at the same time as samples.

Commonly used controls are similar to the controls used in other molecular tests and include:

- A positive control of nucleic acid extraction: this control contains one type of target organism, preferably at lower concentrations. The target organisms can be mixed with plant material that has been previously tested and found negative for the target. In our laboratory, a typical positive control contains 10^3–10^4 target organisms/ml plant extracts, corresponding to the theoretical sensitivity of real-time PCR. For a test to be valid, the positive control for nucleic acid extraction should be positive.
- A negative control of nucleic acid extraction: most commonly this is molecular biology grade water/sample buffer that is processed as a sample, but can also be a plant extract prepared from healthy plants. For a test to be valid, the negative control of nucleic acid extraction should be negative.
- A positive control of real-time PCR: this control contains nucleic acid of the target organism. For a test to be valid, this control should be positive. We have found it very useful to use two positive controls with different concentrations of the target nucleic acids. This allows us to determine real-time PCR amplification efficiency in each run and to follow it over time, thereby providing more data on the robustness of the assay.
- A no-template control (NTC) of real-time PCR: this control contains all the real-time PCR reagents to which molecular biology grade water is added instead of the sample DNA during real-time PCR preparation. Typically, one NTC control is prepared at the beginning of pipetting and one at the end of pipetting. For a test to be valid, these controls should be negative.

Additional controls may further improve the control of the real-time PCR reaction. Those that are most commonly used include:

- Amplification of an endogenous target in all samples: in this control, an additional set of primers and probe are used to amplify a sequence that should be present in the samples provided that the nucleic acid extraction was efficient. For processing mostly plant tissues, a suitable endogenous control example is cytochrome C oxidase (Weller *et al.*, 2000).
- Inhibition control: inhibition in samples can be checked by analysing undiluted and diluted samples. Provided that there is no inhibition, the difference between the Cq values obtained from the undiluted and 1:10 diluted sample should be –3.33. While this can be easily checked for positive samples, it is often inferred from the endogenous control amplification for the negative samples. This assumes that the target amplification and the endogenous control amplification are equally sensitive to inhibitors, which may not always be the case and should, therefore, be determined experimentally.

Additional controls may be needed when performing RT real-time PCR in two steps, e.g. an RNA target as a positive control of the RT step. This requires careful quality control of the RNA templates being investigated and assessment for RT inhibition, and the variability of the RT step can introduce significant errors and uncertainty into the quantification cycles recorded (Bustin, 2012).

3.9 Assay performance and validation

MIQE (minimum information for publication of quantitative real-time PCR experiments) guidelines (Bustin *et al*., 2009) call for the determination and reporting of assay performance characteristics, including PCR amplification efficiency, linear dynamic range, limit of detection and precision. For the field of plant pathogen detection, EPPO has developed guidelines for validation in official plant pathogen diagnostics. The characteristics to be determined for real-time PCR and other methods are defined in the EPPO guideline PM 7/76 (2) (EPPO, 2010) and include analytical sensitivity, analytical specificity, selectivity, repeatability and reproducibility. If considerable variations in the test matrix are expected, the relative insensitivity of a test to variations of the sample material (the matrix effect) needs to be determined; this property is referred to as selectivity. This is often the case in plant pathogen diagnostics as host plants often differ in chemical composition (different varieties, different ages and variable symptom expression). Further guidance on practical determination of these characteristic is given in the appendices to the EPPO guideline PM 7/98 (2) (EPPO, 2014).

Performance characteristics are often determined at two levels: (i) on serial dilutions of purified target DNA; and (ii) on artificially contaminated samples. While the first provides us with a ready approach to assessment of the intrinsic qualities of an assay, the second is more informative as it includes the often significant matrix effect. EPPO guideline PM 7/98 (2) (EPPO, 2014) recommends analysing at least three series of samples artificially contaminated with a range of target concentrations. Both software analysis of raw data and user settings of the baseline and threshold (Fig. 3.1) can influence the determined characteristics of an assay and should be reported. Some assay performance characteristics are described in more detail below.

PCR amplification efficiency, E, is calculated from the slope of the standard curve (Fig. 3.3) using the formula: $E = 10^{(-1/\text{slope})}$, which is usually converted into percentage efficiency, where % efficiency = $((E - 1) \times 100)$. Together with the coefficient of determination (R^2), this reveals how optimal the quantitative real-time PCR (qPCR) assay is and how linear the data are. The **linear dynamic range** is the range over which a detected signal can be quantified, i.e. the range over which the Cq value is linearly related to concentration (Fig. 3.3). The linear dynamic range is often determined visually from the Cq vs concentration plots of a standard curve.

Analytical specificity describes the performance of a test in covering the genetic diversity of the target organism and cross-reactions with non-target organisms. The determination of analytical specificity can include testing of: (i) isolates of the target organism, covering genetic diversity, different geographic origin and hosts; and (ii) a set of non-target organisms, in particular those known to be associated with the sample matrix of interest – these organisms can be characterized or non-characterized, or pathogenic or part of the normal microflora of the host plant. Sample material from the uninfected host plant itself may also be tested. To check for specificity, the DNA is usually tested at relatively high concentrations.

Analytical sensitivity is particularly important for assays aiming to detect low levels of target organisms. In real-time PCR, it describes the lowest target copy number that permits reliable detection according to the EPPO guideline PM 7/98 (EPPO,

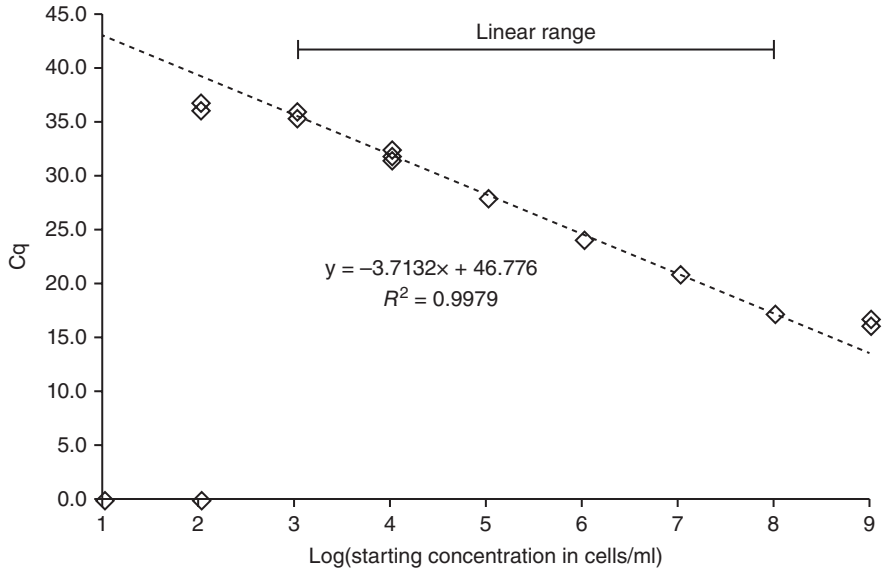


Fig. 3.3. Determination of the linear range and amplification efficiency of a real-time PCR assay. The relationship between the starting amount of the target concentration (cells/ml of plant extract) and the Cq (quantification cycle) values is shown. The linear range in which quantification is accurate is determined visually and in this case extends from log 3 to log 8 cells/ml. The amplification efficiency is calculated from the slope (–3.7132) and in this case corresponds to 86% (see text for details). While 100% amplification efficiency is optimal, efficiencies between 83 and 105% are acceptable for most applications.

2010). In reporting sensitivity, authors rarely define the uncertainty of the determined analytical sensitivity. To determine accurately a concentration at which the probability of detection is, for example, 0.90, high numbers of replicates need to be analysed (Navidi *et al.,* 1992) and this is rarely practical. A useful approximation can be determined, however, through non-linear modelling of a limited amount of data (Burns and Valdivia, 2008; Dreo *et al.*, 2012).

It should be noted that as well as other characteristics of real-time PCR, the analytical sensitivity refers to a specific set of test parameters that should be stringently defined and standardized, including: (i) the brand of PCR reagents (in particular the DNA polymerase); (ii) the PCR cycling conditions; (iii) the inherent characteristics of the real-time PCR assay; (iv) the type and amount of plant material tested; (v) the DNA target copy number; (vi) the DNA extraction and purification method (input into DNA extraction, dilution and/or concentration steps, volume of eluted DNA); (vii) the volume of sample analysed in real-time PCR; (viii) the total reaction volume; and (ix) the number of reactions performed. The reported sensitivity encompasses all of these details and changes in any of them may influence performance of real-time PCR and merit the verification of a method's performance.

The sensitivity of real-time PCR is usually comparable to or better than that of conventional PCR, although this very much depends on the assay design and all of the parameters previously mentioned. Theoretically, even a single copy of a target DNA

can be amplified by PCR and thus generate an amplification curve above the fluorescence background. However, at extremely low target concentrations, the solution composition is not homogeneous. The sampling error (Poisson error) determines the lowest concentration of target DNA that can be detected. In practice, ten copies of target DNA per PCR vessel is the lowest concentration that is amplified each time a PCR assay is performed (Vaerman *et al*., 2004). Any PCR template diluted past a certain threshold copy number will display large variations in amplification (Bustin, 2004) that are also due to the Monte Carlo effect, an inherent limitation of PCR amplification from small amounts of any complex template due to differences in amplification efficiency between individual templates in an amplifying DNA population. If the number of molecules of a particular template is limiting, then that template within a complex mixture will have slight and random differences in amplification efficiencies depending upon whether the primers were able to anneal. If these differences occur early in the PCR reaction, large variations in final product concentration can be produced during the exponential phase of the amplification reaction. DNA of lower abundance will be more likely to experience the Monte Carlo effect because the probability of primer annealing is lower (Fig. 3.4).

Issues of determining analytical sensitivity in the field of plant pathology are further complicated by the absence of certified reference material. While samples with defined numbers of bacteria are fairly straightforward to prepare, this task is much more demanding for pathogens such as viruses and phytoplasmas. It is expected that significant progress will be made in this area through the introduction of digital PCR technology, which will enable accurate quantification of target nucleic acids without the need for standard curves (Baker, 2012; Gutiérrez-Aguirre *et al*., 2015).

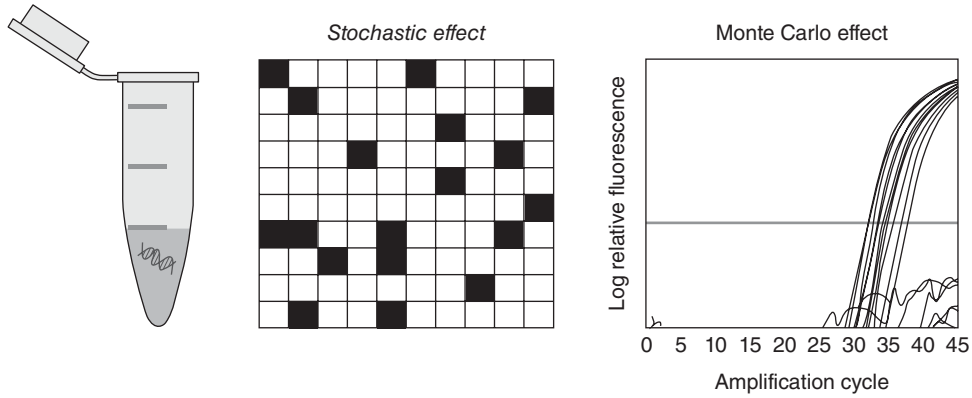


Fig. 3.4. Critical factors affecting the amplification at low target copy numbers. For qualitative detection, the performance of real-time PCR at a low target concentration (left) is crucial. Real-time PCR results are affected at low concentrations of target by the stochastic effect of sampling (centre), leading to positive (black squares) and negative results (white squares) in repeated testing, and by the Monte Carlo effect (right). The Monte Carlo effect describes small differences in the efficiency of primers and probe annealing in the initial amplification cycles, which lead to significant differences in the Cq (quantification cycle) values when the starting number of target sequences is equal but low.

3.10 Qualitative and quantitative real-time PCR

Absolute quantification of the pathogen is not normally a high priority in plant pathology, mainly due to the lack of data on the biological significance of different levels of contamination on disease development. In the case of quarantine diseases, action is decided based on the presence or absence of the organism alone, not its concentration.

The use of PCR simply to detect the presence of a nucleic acid template, rather than to quantify it accurately, is referred to as qualitative PCR, and is widely used in pathogen diagnostics. While this is often considered to be an easier task than quantification, pathogen detection requires an accurate presence/absence answer about the low-end sensitivity of the real-time PCR assay. Consequently, even a qualitative assay should be carefully optimized and should provide information about the assay's performance characteristics that are otherwise commonly associated with quantitative assays (Bustin *et al.*, 2009).

Quantification in real-time PCR relies on the comparison through linear regression of the Cq values obtained for a sample with those obtained for a standard curve, i.e. a group of samples containing known target concentrations. The accuracy of such quantification depends on the efficiency of the PCR amplification and the background fluorescence. The crucial requirement for an accurate quantification is that the amplification efficiency should be the same for samples and for the standard curve.

3.11 Multiplex real-time PCR

For probe-based real-time PCR, more than one assay can be carried out in a single tube, as probes can be labelled with different distinguishable reporter dyes, which also allows the use of appropriate controls in a single reaction. Efficient multiplexing requires the presentation of evidence demonstrating that the accurate quantification of multiple targets in a single tube is not impaired, i.e. that assay efficiency and the limit of detection (LOD) are the same as when the assays are run in simplex fashion. This concern is of particular importance when targets of appreciably lower abundance are co-amplified with highly abundant targets (Bustin *et al.*, 2009). Ishii *et al.* (2007) reported that the proportion of DNA polymerase to PCR amplicon seemed to be critical for minimizing the inhibitory effect of the PCR amplicon accumulated in later cycles of PCR on DNA polymerase activity. They observed that, contrary to expectations, results were better with lower concentrations of primers and probes.

3.12 Confirming the real-time PCR result

As a very sensitive method, real-time PCR is able to detect very low target concentrations. As such, few methods are suitable for the confirmation of a positive result. Sequencing is of limited use because primers and probe cover the larger part of the whole amplicon and sequencing does not provide any new information. To improve the reliability of detection, the use has been proposed of real-time PCR assays targeting different regions of the target genome in combination (see, for example, Gutiérrez-Aguirre *et al.*, 2009; Dreo *et al.*, 2012).

Protocols

Contents

Protocol 3.1

TaqMan real-time PCR detection of *Erwinia amylovora* using a portable SmartCycler® instrument (Cepheid)

This real-time PCR is an example of a protocol originally developed for detection of *E. amylovora* (the fire blight pathogen) using 7900HT (Life Technologies) that was optimized for a faster amplification on a portable SmartCycler® instrument (Cepheid) and the detection of lower target concentrations. Further data on the performance of the assay on both instruments can be found in Dreo *et al.* (2012).

Materials

Samples and controls

- Extracted DNAs of samples and controls (see Notes 1 and 2)
- qPCR (quantitative real-time PCR) reagents
- TaqMan® Fast Universal PCR Master Mix (Life Technologies)
- Primers and probes for AmsC assay labelled with FAM (6-carboxyfluorescein) and BHQ1 (Black Hole Quencher 1) and purified to high-purity, salt-free standard and HPLC (high-performance liquid chromatography), respectively. Primer and probe sequences are as reported in Pirc *et al.* (2009)
- Molecular biology grade water (for no-template controls, NTCs)

Equipment

In addition to general laboratory equipment, the following are required:

- SmartCycler® real-time thermal cycler with corresponding proprietary software
- SmartCycler® reaction tubes (25 µl)

- Centrifuge for SmartCycler® reaction tubes
- Two UV chambers for separate preparation of the reaction mixtures and addition of sample DNAs to reaction mixtures

Method

Setting up plates

1. Determine the number of reactions and remember to include controls.
2. Calculate the volume of components needed for 25 µl reactions, taking into account that the sample volume is 5 µl. Include 10% extra volume for loss during pipetting and possible errors. Each reaction mix should contain the following components at the final concentration indicated:

2× master mix	1×
Forward primer	500 nM
Reverse primer	300 nM
Probe	100 nM
Nuclease-free water	to make up to 20 µl for each reaction.

3. In the first UV chamber, prepare the reaction master mix containing primers and probes. Thaw primers and probe working solutions and allow them to equilibrate to room temperature. Mix all reagents well and spin them down. Aliquot the reaction mixture into tubes (20 µl/tube).
4. In a separate UV chamber, add sample DNAs. By closing the tubes, the sample DNA is pushed into the rectangular reaction compartment of each tube.
5. Briefly centrifuge the tubes and insert them into the SmartCycler®.

Performing real-time PCR

1. Switch on the thermal cycler in advance, according to the manufacturer's instructions, to warm it up and to stabilize temperature and light sources.
2. Run real-time PCR using the following two-step amplification protocol (consult the instrument manual for specific instructions):

Initial denaturation	20 s, 95°C
45 cycles of denaturing and annealing/extension:	
Denaturation	1 s, 95°C
Annealing and extension	20 s, 62°C
Temperature ramping in all stages:	± 3°C

Measure fluorescence in the FAM channel, e.g. FCTC25 during primers and probe annealing step.

Analysing and interpreting the data

1. Analyse the data (see instrument manual for detailed instructions about document set-up, baseline and threshold settings; also Note 3).
2. The test is valid if all controls have the expected result. If a sigmoidal amplification curve is obtained in a sample, increased above the threshold, a sample is considered positive.

Notes

1. Total DNA was extracted using QuickPick™ SML Plant DNA (Bio-Nobile) (Pirc *et al.*, 2009).

2. The following controls are advised: negative isolation control (NIC; molecular biology grade water), positive isolation control containing low target concentration (10^3–10^4 cells/ml of plant extracts), positive amplification control (PAC; *E. amylovora* DNA) and NTCs. In our experience, samples of fire blight hosts prepared as for latent testing according to the PM 7/20 (2) guidelines (EPPO, 2013a) do not contain enough plant material to allow for reliable cytochrome oxidase (COX) amplification. Amplification of COX as an endogenous control is therefore not recommended.

3. The following settings were used for data analysis in Dreo *et al.* (2012): secondary derivative analysis, site-based auto threshold, background subtraction (on 5–10 cycles), standard deviation = 1.

Protocol 3.2

One-step RT-qPCR (reverse transcription quantitative real-time PCR) for detection of *Pepino mosaic virus* (PepMV) in tomato seeds and in water using TaqMan-based chemistry

Materials

Samples and controls

- Extracted RNAs of samples (see Notes 1 and 2)
- Negative isolation control (NIC): extraction method control prepared from clean extraction buffer (see Note 2)
- Negative amplification control (NAC1, NAC2; see Note 3): nuclease-free water – the same as used for the RT-qPCR master mix preparation (see below)
- Positive amplification control (PAC): RNA from a known PepMV infected sample (see Notes 2 and 4)

RT-qPCR reagents

- AgPath-ID™ One-Step RT-PCR Reagents (Ambion): 2× RT-PCR buffer, 25× RT-PCR enzyme mix, nuclease-free water (see Notes 5, 6 and 7)
- Primers and probes (see Note 8) as reported in Ling *et al.* (2007) for the detection of PepMV (see Note 9). Primers and probes used as a control for extraction of seed (see Note 10) or water samples (see Note 11)

Equipment

In addition to general laboratory equipment, the following are required.

- qPCR reaction plates, e.g. 384 well optical reaction plates (Applied Biosystems) (see Notes 5 and 12)
- Optical adhesive covers (e.g. Applied Biosystems)

- Centrifuge with a rotor for PCR reaction plates
- Real-time thermal cycler, e.g. ABI PRISM® 7900 HT Sequence Detection System (Applied Biosystems) (see Note 5)
- Software (e.g. SDS 2.3, Applied Biosystems) for fluorescence acquisition and calculation of threshold cycles (quantification cycles, Cq) (see Note 5)
- Two UV chambers (see Notes 13 and 14)

Method

Setting up plates

1. Arrange the samples in the reaction plate according to the sample number and amplicon (PepMV and COX (cytochrome oxidase)/LUC (luciferase); see Notes 10 and 11). For each primer set, include the NIC, NAC and PAC as a part of the sample group.
2. Calculate the volume of components needed for the assay master mixes. Include sufficient reagents for the number of samples and controls to be tested in triplicate (or at least in duplicate). Include 10% extra volume for loss during pipetting and possible errors. Each reaction mix should contain the following components at the final concentration shown:

2× RT-PCR buffer	1×
Primers and probe	concn of primers/probe (nM) for PepMV 200/400, for COX 900/200 and for LUC 1000/500
25× RT-PCR enzyme mix	1×
Nuclease-free water	to make up final reaction volume of 8 µl.

3. Remove all reagents from the freezer and allow them to thaw, except for the 25× RT-PCR enzyme mix, which should be kept on ice. Mix all thawed reagents well with the vortexer and spin them briefly in a microfuge.
4. Prepare two test assay master mixes (PepMV and COX or LUC) in the individual tubes. For each mix, the reagents are added in the quantity calculated in the previous step. After completion, mix all reagents well with the vortexer and spin them briefly in a microfuge.
5. Take a 384 well qPCR plate. Fill the appropriate number of wells with 8 µl of each master mix according to the plate plan (see Note 15).

6. Add 2 µl of sample and control RNA into each test well (see Note 2). Always start by adding water for the first NAC (NAC1), then continue with RNA samples and controls. Finish with water for the last NAC (NAC2). Using this order of RNA addition facilitates the identification of any possible source of contamination (see Note 15).
7. After adding RNA, the qPCR plate should be covered with an optical adhesive cover applied using a plastic applicator (see Note 16).
8. Centrifuge the plate for 1 min at 1000 × g to ensure that both the reaction mix and the RNA are collected at the bottom of the wells.

Performing the RT-qPCR run

1. Switch on the thermal cycler in advance, according to the manufacturer's instructions, to warm it up and to stabilize temperature and light sources.
2. Transfer the qPCR plate to the real-time thermal cycler and run RT-qPCR using the following cycling programme (consult the instrument manual for specific instructions; see Note 17):

   1 cycle:
   10 min at 48°C (reverse transcription)
   10 min at 95°C (reverse transcriptase inactivation, initial denaturation)
   45 cycles (amplification):
   15 s at 95°C (denaturation)
   1 min at 60°C (annealing and extension).

Analysing data

1. Analyse and export raw data (see instrument manual for detailed instructions about document set-up, baseline and threshold settings) (see Note 18).
2. Import data into Microsoft Excel or equivalent spreadsheet program with statistical features. For each of the replicates of a sample, calculate the average Cq value. If large differences between the replicates are observed, the analysis should be repeated (see Note 19). All samples with an average Cq value lower than the cut-off value should be considered positive for that particular assay (see Note 20).
3. If the target is detected in the NIC, contamination of reagents or samples during RNA extraction should be suspected. The results of sample analysis are therefore not reliable; the analysis is not valid and should be repeated, including the extraction of RNA.
4. NAC1 and NAC2 must not result in a positive qPCR signal; otherwise, contamination during preparation of qPCR should be suspected. The results of the sample analysis are not reliable and the qPCR analyses must be repeated (see Note 21).
5. The presence of target sequence in the PAC must be confirmed. If the target is not detected in the PAC, the analysis is not valid and errors in the qPCR reaction should be considered.

6. The Cq of isolated RNA measured by the endogenous (COX) or external (LUC) controls must be in the predicted range (see Note 22). If not, then the analysis is not valid and some errors in the extraction procedure or the presence of inhibitors should be considered.

Notes

1. Total RNA may be extracted from seed samples using the RNeasy Plant Mini Kit (Qiagen) (see Note 5) following the manufacturer's recommendations and guidance given in EPPO (2013b). The method detected as little as one naturally infected seed among 5000 uninfected seeds (Gutiérrez-Aguirre *et al*., 2009).

The QIAamp Viral RNA mini kit (Qiagen) (see Note 5) may be used for RNA isolation from water samples. LUC RNA may be added to the water samples (2 ng per sample) immediately prior to the isolation as an external control (see Note 11). The concentration of PepMV in water samples can be low, so its detection requires an appropriate concentration step. A variety of methods has been used for concentrating plant viruses in water samples (reviewed by Mehle and Ravnikar, 2012).

2. Mix all RNA samples and controls, and spin them briefly in a microfuge before use. If RNA samples and controls were stored frozen, they should be allowed to thaw before mixing.

3. The NAC contains all the reagents used for RT-qPCR, but nuclease-free water is added into the reaction instead of the sample RNA. The NAC should be used for the same amplicon as the sample in duplicate, one at the start of pipetting (NAC1) and one at the end (NAC2) for each amplicon.

4. To make the assay more uniform, we suggest preparing several aliquots of appropriate controls and putting them in the freezer so that a fresh aliquot can be used for every analysis to prevent degradation of RNA by repeated freezing and thawing. It is recommended to perform two PACs in every RT-qPCR run: one with a high and one with a low concentration of target.

5. Use of names of chemicals or equipment in this chapter implies no approval of them to the exclusion of others. This information is given for the convenience of users and does not constitute an endorsement of the products or services. Equivalent products may be used if they can be shown to lead to the same results.

6. The assay has also been successfully performed by using reagents from other manufacturers, including the ABI One-Step RT-PCR Master Mix Reagents Kit, the Bio-Rad One-Step RT-PCR and the Stratagene Brilliant qPCR Core Reagent kit. It has also been performed as a two-step RT-qPCR format (Gutiérrez-Aguirre *et al.*, 2009; EPPO, 2013b).

7. It is recommended to aliquot the water and buffer.

8. Purchase lyophilized oligonucleotides from any commercial source, synthesized at a 25 nmol scale. For these, standard desalting is sufficient and no additional purification is required. Briefly centrifuge the tubes of dried oligonucleotide to ensure that the

contents are at the bottom of the tubes. Resuspend the oligonucleotides in molecular biology grade water to a concentration of 100 µM. Depending on the vendor, the probe may be supplied in 100 µM solutions. Dilute each stock of forward and reverse primers and probes to working concentrations (e.g. 10 µM); prepare aliquots of these working solutions and store them at −20°C. When stored properly and subjected to minimal freeze–thaw cycles, a primer/probe stock may last at least 3 years. Protect the probe from excessive exposure to light (e.g. put the tubes with probes in dark plastic bags/containers) to prevent photobleaching of the fluorescent dyes and evaporation.

9. This assay is targeted at the *TGB2* gene of the PepMV genome and comprises two primer pairs with a single TaqMan probe. It can specifically detect two PepMV genotype groups: Peruvian-European tomato-US1 using one of the pair of primers, and Ch2-US2 using the second pair of primers. When applying both pairs of primers in a single reaction, all PepMV genotypes can be detected, so this assay can be used as a 'universal' PepMV detection assay. If genotype discrimination of PepMV is needed (e.g. for epidemiological studies or studies evaluating the resistance of plants to virus infection), specific primers and probes (Gutiérrez-Aguirre *et al*., 2009) may be used.

10. In order to control the extraction procedure from plant material (in this case tomato seeds), an endogenous control must be included in the assays. This is preferably an assay designed to a conserved gene, which provides confirmation that the extraction from the plant material has been successful, and it should be present in higher copy numbers in the genome in order to increase the sensitivity of the assay. Instead of plant COX, primers and probes adapted from Weller *et al*. (2000) with the following sequence (5′–3′) may be used: forward primer, CGTCGCATTCCAGATTATCCA; reverse primer, CAACTACGGATATATAAGRRCCRRAACTG; and probe, FAM-AGGGCATTCCATCCAGCGTAAGCA-TAMRA or another endogenous control, e.g. 18S rRNA. (A, T, C and G are the basic four DNA nucleotides; FAM is 6-carboxyfluorescein and TAMRA is tetramethylrhodamine.)

11. In order to control the extraction procedure from water samples, an external control must be included in the assays, e.g. LUC RNA (see Note 1); primers and probes for LUC are reported in Toplak *et al.* (2004).

12. Alternatively, 96 well optical reaction plates can be used. If 96 well plates are used, the reaction volume should be adjusted (e.g. to 25 µl).

13. Reaction mixes and the addition of RNA samples should be prepared in separate UV chambers using dedicated laboratory equipment (pipettes, tips, tubes, microcentrifuge tube opener, racks for tubes, laboratory coat, gloves). First, prepare reaction mixes and load them on to the qPCR plate (no RNA/DNA other than primers and probes should be present here); second, add RNA of samples in another UV chamber. Both steps should be done in a separate location from the places used for homogenization of the samples and RNA extraction.

14. In all stages of setting up PCR reactions, the following precautions should be taken in order to avoid contamination of samples and reagents: (i) use only filter tips (aerosol barrier tips) and use a fresh tip for each pipetting step; (ii) close reagent/sample tubes

once the desired volume has been aspirated to the tip; (iii) always wear gloves and change them if you suspect contamination; (iv) use only clean, sterile plastic ware; (v) all tubes should be opened by using a microfuge tube opener.

15. Start with pipetting the master mixes on to the bottom of the wells and continue by adding RNA samples and controls to the side walls of the wells. This facilitates visual checking during the addition of samples if no devices for pipetting navigation (e.g. PlatR™ from BioSistemika) are available in the laboratory. Avoid touching the master mixes inside the tubes with the pipette tip. It is advisable to use multistep pipettors. In this case, use one tip to fill up all wells with the same master mix and one tip for three (two) replicates when adding the RNA of the sample. An automated liquid handling system (e.g. Multiprobe® II PLUS EX from PerkinElmer) may be used for pipetting a large number of RNA samples and master mixes on to the 384 well plates. In that case, the pipetting order of master mixes and RNA may be reversed: RNA is applied first (on to the bottom of the wells), and the plate is next centrifuged to ensure that the RNA is collected at the bottom, and then master mixes are added without touching the well bottoms.

16. Do not touch the optical adhesive cover with fingers.

17. Before starting the run: (i) adjust the reaction volume, if needed; (ii) assign samples to locations on the plate; (iii) assign detectors to the samples (choose the correct filters according to the probes used in amplicons: FAM for PepMV and COX amplicons, and VIC® for the LUC amplicon). The temperature ramp rate should be similar to that of the Applied Biosystems 7900 thermal cyclers. Using faster temperature ramp rates can reduce assay sensitivity.

18. The transformation of the fluorescence signal into Cq data (e.g. Ct, the threshold cycle; Cp, the crossing point), as well as methods for baseline and threshold settings, vary between instrument models. The specific instrument manual must be consulted. When analysing the raw data, it is important to adjust the cycle threshold (Cq) of the amplification plot to within the geometric (exponential) phase of amplification, preferably at the beginning of the geometric phase. In the log view, this is the linear part of the amplification plot. Manually check the multi-component plots for all samples (wells). In wells containing positive samples, the fluorescence of a fluorescent reporter (in this case the FAM dye) should be increasing with the number of cycles. A Cq value where no increase of FAM fluorescence is observed must not be considered as positive.

19. If the Cq values are very high, the use of replicates is especially important to make sure that the reaction is working properly. If data from the replicate reactions vary significantly, the reactions should be repeated, unless high variability in the Cq values is attributed to the stochastic effect of target copy distribution in replicate reactions. This occurs when less than ten copies of the target are present in the reaction (Morrison *et al.*, 1998; Ellison *et al.*, 2006).

20. The cut-off value is determined for each assay in the process of assay validation before the assay is put into routine use. This is the highest Cq value that is still considered to be a positive result. Different guidelines may be followed for the determination

of the cut-off value during validation of the assay (Mehle *et al.*, 2013). The cut-off value should be determined for each assay at the time of the method validation. For the same assay, it may vary from laboratory to laboratory for various reasons (e.g. the qPCR machine and software used). Therefore, each laboratory should revalidate the cut-off values for any new assay prior to its implementation.

21. If contamination of the reagents is suspected, use fresh aliquots of all reaction components. The amplification in the NAC may also be a result of degradation of the primer mix, or of the amplification of non-specific PCR products.

22. The quality and quantity of isolated RNA greatly influences the detection of PepMV. Monitoring the Cq values of endogenous or external controls allows the estimation of both parameters, thereby increasing confidence in the results. The Cq values for LUC/COX in the RNA extraction procedure used in our laboratory for water/seed samples (see Note 1) vary from 12 to 16/18 to 23. Higher Cq values are an indicator of lower amounts of LUC/seed RNA and presumably of lower amounts of PepMV RNA in the sample, which therefore decreases confidence in the results. This is a signal that such samples should be re-extracted and retested. Sometimes, higher Cq values are only an indicator of inhibitors present in the sample, therefore tenfold dilutions of extracted RNAs of samples should be retested.

Protocol 3.3

One-step RT-qPCR (reverse transcription quantitative real-time PCR) for detection and discrimination of *Potato virus Y* (PVY) isolates

Materials

Samples and controls

- Extracted RNAs of samples (see Note 1)
- Negative isolation control (NIC): extraction method control prepared from clean extraction buffer (see Note 2 in Protocol 3.2)
- Negative amplification control (NAC1, NAC2; see Note 3 in Protocol 3.2): nuclease-free water – the same as used for RT-qPCR master mix preparation (see below)
- Positive amplification control (PAC): RNA from a known PVY infected sample (see Notes 2 and 4 in Protocol 3.2)

RT-qPCR reagents

- Brilliant QPCR Core Reagent kit (Stratagene): 10× Core PCR buffer, 5 U/µl SureStart *Taq* DNA polymerase, 50 mM magnesium chloride, 20 mM dNTP (deoxynucleotide triphosphate) mix, 1 mM reference dye (diluted 1:50) (see Note 5 in Protocol 3.2 and Notes 2 and 3 below)
- 200 U/µl M-MLV (Moloney Murine leukemia virus) reverse transcriptase (Promega; diluted 1:100)
- Molecular biology grade water (nucleic acid and nuclease free)
- Primers and probes (see Note 8 in Protocol 3.2) as reported in Kogovšek *et al.* (2008) for detection of all PVY isolates (Univ F and R primer and Univ probe)
- Primers and probes as reported in Kogovšek *et al.* (2008) for discrimination of PVY isolates (see Note 4)

- Primers and probes used as a control for extraction (see Note 10 in Protocol 3.2)

Equipment

See Protocol 3.2

Method

Setting up plates

1. Arrange the samples in the reaction plate according to the sample number and amplicon (see Note 4). For each primer set, include the NIC, NAC and PAC as a part of the sample group.
2. Calculate the volume of components needed for the assay master mixes. Include sufficient reagents for the number of samples and controls to be tested in triplicate (or at least in duplicate). Include 10% extra volume for loss during pipetting and possible errors. Each reaction mix should contain the following components at the final concentration shown:

10× Core PCR buffer	1x
50 mM magnesium chloride	5.5 mM
20 mM dNTP mix	0.8 mM
1 mM reference dye (diluted 1:50)	300 nM (see Note 3)
200 U/µl M-MLV (diluted 1:100)	0.8 U
Primers and probe	concn. of primers/probe (nM) for PVY 300/150, for COX 900/200
5 U/µl SureStart *Taq* DNA polymerase	0.5 U
Nuclease-free water	to make up final reaction volume of 9 µl.

3. Remove all reagents from the freezer and allow them to thaw. Gently mix all thawed reagents and spin them briefly in a microfuge. All reagents should be kept on ice.
4. Prepare test assay master mixes in individual tubes. For each mix, the reagents are added in the quantity calculated in the previous step (Note 2). After completion, gently mix all the reagents and spin them briefly in a microfuge.
5. Take a 384 well qPCR plate. Fill the appropriate number of wells with 9 µl of each master mix according to the plate plan (see Note 15 in Protocol 3.2).

6. Add 1 µl of sample and control RNA into each test well (see Note 2 in Protocol 3.2). Always start by adding water for the first NAC (NAC1), continue with RNA samples and controls. Finish with water for the last NAC (NAC2). Using this order of RNA addition facilitates the identification of any possible source of contamination (see Note 15 in Protocol 3.2).
7. After adding RNA, the qPCR plate should be covered with an optical adhesive cover applied using a plastic applicator (see Note 16 in Protocol 3.2).
8. Centrifuge the plate for 1 min at 1000 × g to ensure that both the reaction mix and the RNA are collected at the bottom of the wells.

Performing the RT-qPCR run

1. Switch on the thermal cycler in advance, according to the manufacturer's instructions, to warm it up and to stabilize temperature and light sources.
2. Transfer the qPCR plate to the real-time thermal cycler and run the RT-qPCR using the following cycling programme (consult the instrument manual for specific instructions; see Note 17 in Protocol 3.2):

   1 cycle:
   30 min at 48°C (reverse transcription)
   10 min at 95°C (reverse transcriptase inactivation, initial denaturation)
   40 cycles (amplification):
   15 s at 95°C (denaturation)
   1 min at 60°C (annealing and extension).

Analysing data

See Protocol 3.2 and Protocol 3.3, Table 1.

Protocol 3.3, Table 1. Interpretation of RT-qPCR results in order to identify the PVY strain (PVY^{N}, PVY^{NTN}, PVY^{O}) present in a sample.

Amplicon				
PVY-Uni	**PVY-N**	**PVY-NTN**	**PVY-O**	**Conclusion**
neg[a]	nt[a]	nt	nt	PVY not present
pos[a]	pos	neg	neg	PVY^{N} present
pos	neg	pos	neg	PVY^{NTN} present
pos	neg	pos	pos	PVY^{O} or PVY^{O} and PVY^{NTN} present (see Note 4)

[a]neg, negative result; nt, not tested; pos, positive result.

Notes

1. Total RNA may be extracted from plant material (leaves, tubers) using the RNeasy Plant Mini Kit (Qiagen, USA) (see Note 5 in Protocol 3.2) following the manufacturer's recommendations. In routine testing, immune-capture and post-ELISA virus release may be coupled with RT-qPCR (Kogovšek *et al.*, 2008).

2. The assay has been successfully performed by also using reagents from other manufacturers including Ambion AgPath-ID™ One-Step RT-PCR Reagents and as two-step RT-qPCR format (Kogovšek *et al.*, 2008).

3. Prepare fresh dilutions of the reference dye according to the manufacturer's recommendations (e.g., 1:50 if ABI PRISM® 7900 HT Sequence Detection System is used) prior to setting up the reactions, and keep tube containing the reference dye protected from light. The shelf life of diluted reference dye is limited and it should be used on the day of dilution. The addition of the reference dye is not required for all instruments.

4. RT-qPCR with PVY-Uni and COX (see Note 10 in Protocol 3.2) amplicons is used to screen for the presence of all PVY isolates in samples. Thereafter, RT-qPCR with the amplicons PVY-NTN, PVY-N and PVY-O enables differentiation between potentially tuber necrotic PVY^{NTN} isolates and standard PVY^{N} isolates, and detection of the presence of PVY^{O} isolates. PVY^{O} isolates give positive results with PVY-NTN and PVY-O amplicons. It is not possible to exclude mixed infection with PVY^{NTN} without additional analysis of the samples by ELISA (Kogovšek *et al.*, 2008).

References

Baker, M. (2012) Digital PCR hits its stride. *Nature Methods* 9, 541–544.

Bentley, H.A., Belloni, D.R. and Tsongalis, G.J. (2005) Parameters involved in the conversion of real-time PCR assays from the ABI prism 7700 to the Cepheid SmartCycler® II. *Clinical Biochemistry* 38, 183–186.

Bester, R., Jooste, A.E.C., Maree, H.J. and Burger, J.T. (2012) Real-time RT-PCR high-resolution melting curve analysis and multiplex RT-PCR to detect and differentiate grapevine leafroll-associated virus 3 variant groups I, II, III and VI. *Virology Journal* 9, 219.

Buh Gašparič, M., Cankar, K., Žel, J. and Gruden, K. (2008) Comparison of different real-time PCR chemistries and their suitability for detection and quantification of genetically modified organisms. *BMC Biotechnology* 8, 26.

Burns, M. and Valdivia, H. (2008) Modelling the limit of detection in real-time quantitative PCR. *European Food Research and Technology* 226, 1513–1524.

Bustin, S.A. (ed.) (2004) *A–Z of Quantitative PCR*. IUL Biotechnology Series No. 5, International University Line, La Jolla, California.

Bustin, S.A. (2012) *Expert Guides. Definitive qPCR: Basic Principles, v1.2*. Stephen Bustin, USA. Available at: http://www.qpcrexpert.com/practical-guide.html (accessed 27 October 2015).

Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L. *et al.* (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry* 55, 611–622.

Dreo, T., Pirc, M. and Ravnikar, M. (2012) Real-time PCR, a method fit for detection and quantification of *Erwinia amylovora*. *Trees* 26, 165–178.

Ellison, S.L.R., English, C.A., Burns, M.J. and Keer, J.T. (2006) Routes to improving the reliability of low level DNA analysis using real-time PCR. *BMC Biotechnology* 6, 33.

EPPO (2010) PM 7/76 (2): use of EPPO diagnostic protocols. *EPPO Bulletin* 40, 350–352.

EPPO (2013a) PM 7/20 (2): *Erwinia amylovora*. *EPPO Bulletin* 43, 21–45.

EPPO (2013b) PM 7/113 (1): *Pepino mosaic virus*. *EPPO Bulletin* 43, 94–104.

EPPO (2014) PM 7/98 (2): specific requirements for laboratories preparing accreditation for a plant pest diagnostic activity. *EPPO Bulletin* 44, 117–147.

Gabriel, D.W., Allen, C., Schell, M., Denny, T.P., Greenberg, J.T., Duan, Y.P., Flores-Cruz, Z., Huang, Q., Clifford, J.M., Presting, G. *et al.* (2006) Identification of open reading frames unique to a select agent: *Ralstonia solanacearum* race 3 biovar 2. *Molecular-Plant Microbe Interactions* 19, 69–79.

Gori, A., Cerboneschi, M. and Tegli, S. (2012) High-resolution melting analysis as a powerful tool to discriminate and genotype *Pseudomonas savastanoi* pathovars and strains. *PLoS ONE* 7(1), e30199.

Gutiérrez-Aguirre, I., Mehle, N., Delic, D., Gruden, K., Mumford, R. and Ravnikar, M. (2009) Real-time quantitative PCR based sensitive detection and genotype discrimination of *Pepino mosaic virus*. *Journal of Virology Methods* 162, 46–55.

Gutiérrez-Aguirre, I., Rački, N., Dreo, T. and Ravnikar, M. (2015) Droplet digital PCR for absolute quantification of pathogens. *Methods in Molecular Biology* 1302, 331–347.

Ishii, T., Sootome, H. and Yamashita, K. (2007) Practical evaluation of universal conditions for four-plex quantitative PCR. *Analytical and Bioanalytical Chemistry* 388, 271–278.

Juskowiak, B. (2011) Nucleic acid-based fluorescent probes and their analytical potential. *Analytical and Bioanalytical Chemistry* 399, 3157–3176.

Kogovšek, P., Gow, L., Pompe Novak, M., Gruden, K., Foster, G.D., Boonham, N. and Ravnikar, M. (2008) Single-step RT real-time PCR for sensitive detection and discrimination of *Potato virus Y* isolates. *Journal of Virological Methods* 149, 1–11.

Koshkin, A.A., Singh, S.K., Nielsen, P., Rajwanshi, V.K., Kumar, R., Meldgaard, M., Olsen, C.E. and Wengel, J. (1998) LNA (locked nucleic acids): synthesis of the adenine, cytosine,

guanine, 5-methylcytosine, thymine and uracil bicyclonucleoside monomers, oligomerisation, and unprecedented nucleic acid recognition. *Tetrahedron* 54, 3607–3630.

Kutyavin, I.V., Afonina, I.A., Mills, A., Gorn, V.V., Lukhtanov, E.A., Belousov, E.S., Singer, M.J., Walburger, D.K., Lokhov, S.G., Gall, A.A. *et al*. (2000) 3′-Minor groove binder-DNA probes increase sequence specificity at PCR extension temperatures. *Nucleic Acids Research* 28, 655–661.

Ling, K.S., Wechter, W.P. and Jordan, R. (2007) Development of a one-step immunocapture real-time TaqMan RT-PCR assay for the broad spectrum detection of *Pepino mosaic virus*. *Journal of Virological Methods* 144, 65–72.

Mackay, I.M., Arden, K.E. and Nitsche, A. (2002) Real-time PCR in virology. *Nucleic Acids Research* 30, 1292–1305.

Mehle, N. and Ravnikar, M. (2012) Plant viruses in aqueous environment: survival, water mediated transmission and detection. *Water Research* 46, 4902–4917.

Mehle, N., Prezelj, N., Hren, M., Boben, J., Gruden, K., Ravnikar, M. and Dermastia, M. (2013) A real-time PCR detection system for the bois noir and flavescence dorée phytoplasmas and quantification of the target DNA. *Methods in Molecular Biology* 938, 253–268.

Morrison, T.B., Weis, J.J. and Wittwer, C.T. (1998) Quantification of low-copy transcripts by continuous SYBR Green I monitoring during amplification. *BioTechniques* 24, 954–958, 960, 962.

Navidi, W., Arnheim, N. and Waterman, M.S. (1992) A multiple-tubes approach for accurate genotyping of very small DNA samples by using PCR: statistical considerations. *American Journal of Human Genetics* 50, 347–359.

Palmano, S., Mulholland, V., Kenyon, D., Saddler, G.S. and Jeffries, C. (2015) Diagnosis of phytoplasmas by real-time PCR using locked nucleic acid (LNA) probes. In: Lacomme, C. (ed.) *Plant Pathology*. Springer New York, pp. 113–122.

Pirc, M., Ravnikar, M., Tomlinson, J. and Dreo, T. (2009) Improved fireblight diagnostics using quantitative real-time PCR detection of *Erwinia amylovora* chromosomal DNA. *Plant Pathology* 58, 872–881.

Pritchard, L., Humphris, S., Saddler, G.S., Parkinson, N.M., Bertrand, V., Elphinstone, J.G. and Toth, I.K. (2013) Detection of phytopathogens of the genus *Dickeya* using a PCR primer prediction pipeline for draft bacterial genome sequences. *Plant Pathology* 62, 587–596.

Ruijter, J.M., Pfaffl, M.W., Zhao, S., Spiess, A.N., Boggy, G., Blom, J., Rutledge, R.G., Sisti, D., Lievens, A., De Preter, K. *et al*. (2013) Evaluation of qPCR curve analysis methods for reliable biomarker discovery: bias, resolution, precision, and implications. *Methods* 59, 32–46.

Szemes, M., Klerks, M.M., van den Heuvel, J.F.J.M. and Schoen, C.D. (2002) Development of a multiplex AmpliDet RNA assay for simultaneous detection and typing of *Potato virus Y* isolates. *Journal of Virological Methods* 100, 83–96.

Toplak, N., Okršlar, V., Stanič, D., Gruden, K. and Žel, J. (2004) A high-throughput method for quantifying transgene expression in transformed plants with real-time PCR analysis. *Plant Molecular Biology Reporter* 22, 237–250.

Trombley Hall, A., McKay Zovanyi, A., Christensen, D.R., Koehler, J.W. and Devins Minogue, T. (2013) Evaluation of inhibitor-resistant real-time PCR methods for diagnostics in clinical and environmental samples. *PLoS ONE* 8(9), e73845.

Tyagi, S. and Kramer, F.R. (1996) Molecular beacons: probes that fluoresce upon hybridization. *Nature Biotechnology* 14, 303–308.

Tyagi, S. and Kramer, F.R. (2012) Molecular beacons in diagnostics. *F1000 Medical Reports* 4, 10.

Vaerman, J.L., Saussoy, P. and Ingargiola, I. (2004) Evaluation of real-time PCR data. *Journal of Biological Regulators and Homeostatic Agents* 18, 212–214.

Weller, S.A., Elphinstone, J.G., Smith, N.C., Boonham, N. and Stead, D.E. (2000) Detection of *Ralstonia solanacearum* strains with a quantitative, multiplex, real-time, fluorogenic PCR (TaqMan) assay. *Applied and Environmental Microbiology* 66, 2853–2858.

4 Loop-mediated isothermal amplification (LAMP)

Jenny Tomlinson*

Fera, Sand Hutton, York, UK

4.1 Introduction

Detection methods based on the polymerase chain reaction (PCR) are well established and allow sensitive and specific detection of pests and pathogens, typically within a few hours. However, PCR requires accurate thermal cycling, and the instrumentation required for this is relatively complex and expensive, and so its use is to a large extent limited in laboratory facilities. PCR amplification products are conventionally detected using methods such as gel electrophoresis. This approach is inexpensive and flexible, but is relatively time-consuming and not ideally suited to high-throughput applications. Amplification by PCR can be monitored in real time by using fluorescence; while this reduces the time required and enables a higher throughput, relatively sophisticated equipment is required to perform concurrent thermal cycling and fluorescence measurement. Amplification methods that work under isothermal conditions have been developed as alternatives to PCR. Because these methods do not require thermal cycling, they can be carried out using simpler equipment and, as a result, they are potentially more accessible for use in locations where complex and/or costly instrumentation is not available.

In each cycle of denaturation and annealing of a PCR reaction, primers bind to the template nucleic acid and are extended by DNA polymerase, resulting in exponential amplification of the target sequence. A mechanism for isothermal amplification therefore requires a means of generating single-stranded primer-binding sites to allow replication to proceed without repeated rounds of thermal denaturation. Some methods require an initial denaturation step at the start of the reaction before amplification proceeds at a single temperature, but these methods are still considered to

*E-mail: jenny.tomlinson@fera.co.uk

be isothermal as the primary reaction occurs without thermal cycling. There are a range of possible approaches to the generation of primer-binding sites without thermal denaturation in order to achieve amplification of target nucleic acid (Gill and Ghaemi, 2008). These include non-thermal mechanisms for denaturation as well as methods based on strand displacement or transcription of RNA. Particularly efficient amplification can be achieved via the formation of a secondary structure containing single-stranded regions, such as loops, to which primers can bind without thermal denaturation and reannealing.

4.2 Loop-mediated isothermal amplification (LAMP)

As already noted, one approach to isothermal amplification is to generate products containing self-complementary regions that form single-stranded loops to which primers can bind. Loop-mediated isothermal amplification (LAMP) is the most common method using this approach and was first described by Notomi *et al.* (2000). LAMP uses two pairs of primers (internal and external) and a DNA polymerase with strand-displacing activity to generate amplification products containing loops (Fig. 4.1). The internal primers target two different binding sites of opposite orientations such that the extension product contains self-complementarity and forms a loop. In the early stages of the reaction, this extension product is displaced from the template strand by extension of the adjacent external primer, but this is not required later on in the reaction. The amplification product consists of structures composed of differing numbers of alternately oriented repeats of the target sequence, resulting in a ladder-like appearance when visualized by gel electrophoresis (see Fig. 4.2), in contrast to PCR and other amplification methods, which typically generate products consisting of a single structure of a defined size. In the initial description of the LAMP mechanism (Notomi *et al.*, 2000), an initial denaturation step was used to allow the primers to bind before the reaction proceeded isothermally, but Nagamine *et al.* (2001) demonstrated that amplification could occur efficiently without this step, thus making the method completely isothermal. Amplification can be accelerated by the addition of loop primers (Nagamine *et al.*, 2002); these bind to loops in the amplification product that are not

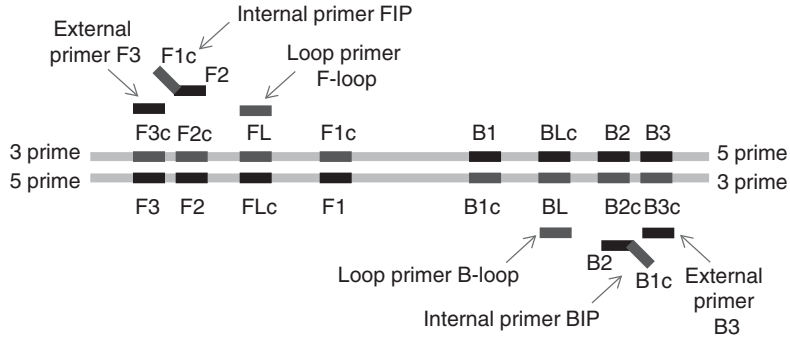


Fig. 4.1. Primers used for loop-mediated isothermal amplification (LAMP) of DNA. Key: FIP and BIP, forward and backward internal primers; F-loop and B-loop, forward and backward loop primers; F3 and B3, forward and backward external primers.

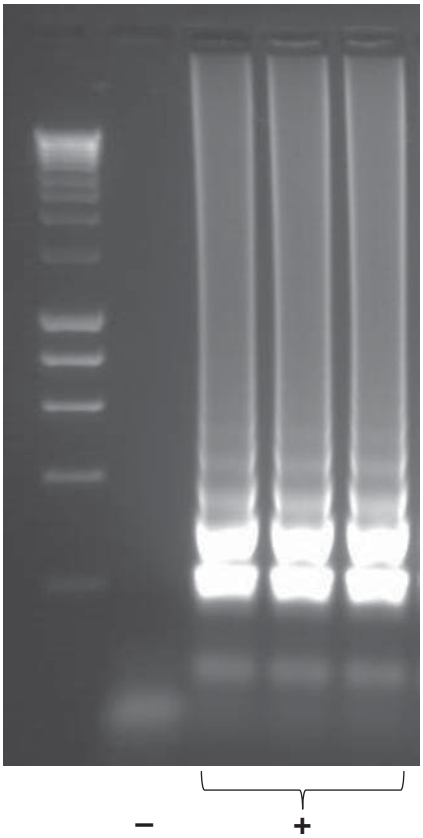


Fig. 4.2. LAMP product visualized by agarose gel electrophoresis.

in the correct orientation to bind the internal primers. Loop primers are not essential for the amplification mechanism to proceed, so LAMP assays may be designed which lack loop primers, or which include only one loop primer; however, assays with two loop primers typically amplify target DNA in a shorter period of time. Gandelman *et al.* (2011) described the use of 'stem' primers to increase the speed of amplification and to allow additional options in LAMP assay design.

The LAMP reaction mechanism relies on the strand displacement activity of the DNA polymerase used and, as a result, *Bst* polymerase (derived from *Bacillus stearothermophilus*), which has an optimal reaction temperature of 65°C, is commonly used. LAMP reactions are typically carried out at 65°C (or in the range 63–66°C) for up to 60 min. More recently developed reagents using an alternative enzyme, *Gsp* M2.0 (an engineered variant of GspM LF DNA polymerase, isolated from *Geobacillus* sp. M; OptiGene), allow efficient amplification in significantly reduced reaction times of 30 min or less. RNA targets can be detected by LAMP by also adding reverse transcriptase to the reaction. Thermostable reverse transcriptase, such as ThermoScript (Life Technologies), derived from cloned *Avian* myeloblastosis virus (AMV), is active at 65°C, so one-tube reverse transcriptase-LAMP (RT-LAMP) can be carried out using the same reaction conditions as LAMP for DNA amplification (Nie, 2005; Tomlinson *et al.*, 2013).

The use of 4–6 primers in LAMP confers a high degree of specificity if the primers are designed at the locations of mismatches between the target and related non-target sequences. The sensitivity of LAMP assays typically exceeds that of conventional PCR designed to target the same sequence (Fukuta *et al.*, 2003; Zhang *et al.*, 2011), and can approach that of comparable real-time PCR assays (Tomlinson *et al.*, 2007). The primary advantages of LAMP over conventional PCR-based approaches are, therefore, the speed of amplification and the potential to perform the reactions using very simple equipment such as a water bath or a heated block.

LAMP-based methods have been developed for a range of plant pathogens and pests (Tomlinson and Boonham, 2008), including viruses and viroids (e.g. Fukuta *et al.*, 2003; Boubourakas *et al.*, 2009), fungi and oomycetes (e.g. Tomlinson *et al.*, 2007), bacterial pathogens, including phytoplasmas (e.g. Tomlinson *et al.*, 2010b; Lenarčič *et al.*, 2014), nematodes (e.g. Kikuchi *et al.*, 2009) and insect pests (e.g. Arif *et al.*, 2012). In most cases, the rationale for the development of LAMP-based tests for these targets included improvements in speed and simplicity in comparison with conventional methods and, in particular, the lower equipment costs relative to conventional PCR.

One feature of LAMP as a nucleic acid amplification method that is potentially useful in the detection of plant pathogens is the greater tolerance to inhibitors of the DNA polymerase used for LAMP in comparison with those used for conventional PCR (typically *Taq* polymerase). As mentioned in Chapter 2, plant material and other matrices relevant to plant pathology, such as soil, can contain high levels of inhibitory substances (such as polysaccharides, polyphenolics and humic acid) which are known to adversely affect PCR (Kaneko *et al.*, 2007; Tani *et al.*, 2007), and these need either to be removed in the nucleic acid extraction process, or their effects counteracted by the use of PCR additives. By using enzymes that are tolerant of potentially inhibitory substances, LAMP enables the use of more rudimentary extraction methods, which are compatible with rapid, simplified testing methods.

4.3 LAMP detection methods

There are numerous ways in which the products of a LAMP reaction (or other amplification reaction) can be detected; these can be categorized as end-point methods in which the product is detected after completion of the reaction, and real-time methods in which amplification is monitored during the reaction.

LAMP amplification products can be visualized by gel electrophoresis, in the same way as the products of other nucleic acid amplification methods (see Fig. 4.2). LAMP is highly efficient, and the amount of product generated also allows the use of other end-point detection methods, which would be insufficiently sensitive to detect the smaller amounts of product generated by amplification methods such as PCR. A particularly simple method for the detection of amplification by LAMP is the observation of turbidity caused by the precipitation of magnesium pyrophosphate, a by-product of polymerization (Mori *et al.*, 2001). Turbidity can be measured at the end of the reaction using relatively simple and inexpensive equipment; the white precipitate can be visible to the unaided eye, but detection can be less robust using this method (Wastling *et al.*, 2010).

In an alternative approach, the fluorescent metal indicator calcein can be added to the reaction with Mn^{2+}. As pyrophosphate is generated, it binds to the manganese ions, releasing calcein and resulting in fluorescence (Tomita *et al.*, 2008).

Fluorescent dyes such as SYBR Green, which intercalate between the strands of double-stranded DNA (dsDNA), can also to be used as indicators for end-point detection of amplification by LAMP, and result in a clear colour change (typically from orange to yellow/green) (Iwamoto *et al.*, 2003). The concentrations of intercalating dye typically required to produce an unambiguous visible colour change are inhibitory to amplification, so the dye must be added to the reaction tube after amplification is completed. This is a significant disadvantage of the intercalating dye method (and of other open-tube methods such as electrophoresis) due to the risk of carry-over contamination, and if using this approach, pre- and post-amplification steps should be kept separate at all times. Closed-tube detection methods are generally preferable, including the use of calcein/Mn^{2+} as described above. An alternative closed-tube colour change reaction uses the metal ion indicator hydroxy naphthol blue (HNB) to produce a violet/blue colour change at concentrations that do not adversely affect amplification (Goto *et al.*, 2009).

Detection methods based on turbidity, and the majority of colour change methods, are non-specific in that they give an indication of the total amount of product (or by-product) generated. An alternative end-point detection method allows the detection of specific LAMP products, with the potential to resolve the mixed products of multiplex amplification. This approach involves the detection of labels incorporated into the amplification products using labelled primers or a combination of primers and probes. A pair of labels can be used to dual label specific amplification products, and the dual-labelled products can be detected, for example, by using a sandwich immunoassay in a lateral flow device (LFD) format (Kiatpathomchai *et al.*, 2008; Tomlinson *et al.*, 2010a). Each assay can be assigned a different pair of labels, such that one LAMP product can be discriminated from another. Opening the reaction tube to transfer the amplification product to the LFD poses a contamination risk, as discussed above, but there is the potential to incorporate LFDs into a sealed cassette to avoid the need for open-tube manipulations.

End-point detection methods have the advantage of requiring only minimal equipment (for example, a water bath or simple heated block for incubation of the LAMP reaction tubes). However, real-time detection has the advantage of using a closed-tube system, and this can outweigh the disadvantage of requiring specialized instrumentation. Turbidity resulting from pyrophosphate production can be monitored in real time (Mori *et al.*, 2004), and this can be performed using relatively simple instrumentation. Real-time fluorescence detection is increasingly a preferred approach (Tomlinson *et al.*, 2010c; Bühlmann *et al.*, 2013a,b; Lenarčič *et al.*, 2014), despite the requirement for an instrument for incubating reaction tubes at a constant temperature with concurrent fluorescence monitoring. Reasons for this preference include the greatly reduced risk of contamination and the ability to include melt curve analysis for increased robustness of results. Instruments intended for real-time PCR can also be used for real-time LAMP, as the same fluorescent intercalating dyes can be used with both amplification methods. Less costly instruments have been developed specifically for real-time LAMP, allowing

the end user to take advantage of the greater simplicity and shorter run times of LAMP. One example is the Genie II instrument (OptiGene), which can be used to run 16 LAMP reactions in one run with a typical reaction time of 30–40 min. Other low-cost instruments for isothermal amplification include the T-16 and Twista instruments (TwistDx). Using an intercalating dye for real-time fluorescence monitoring during amplification also allows the specificity of the amplification product to be confirmed by performing melting or annealing analysis at the end of the reaction (Fig. 4.3). On the Genie II instrument, this is typically achieved by subjecting the post-amplification reaction to a gradual decrease in temperature during which the fluorescence signal is monitored. The strands of the amplification product will anneal together at a specific temperature, causing a rapid increase in fluorescence, which is consistent and reproducible for that assay. This allows specific amplification products to be distinguished from any non-specific products or artefacts such as primer dimers.

In conclusion, the efficiency of amplification achieved using LAMP distinguishes this method from many other amplification techniques but also constitutes a significant potential disadvantage of the method, as the potential of carry-over contamination with amplification products is substantially higher than for other methods. In selecting a detection method, it is, therefore, necessary to be mindful of this disadvantage in conjunction with other considerations such as cost and throughput. While the risk of carry-over contamination is the most likely practical drawback to the use of LAMP for plant pathogen detection, relatively simple steps can be taken to mitigate this risk. Stringent separation of all pre- and post-amplification manipulations can be effective in virtually eliminating carry-over contamination provided that the required facilities for this are available.

4.4 LAMP primer design

Approaches to the design of LAMP primers for the detection of pathogens and pests are the same as for PCR-based methods. In summary, species-specific primers can be designed with reference to alignments of sequence from target and non-target organisms, and specificity will result from locating the extendable (3′) end of primers at the positions of mismatched bases. Conversely, inclusive detection of sequence variants can be

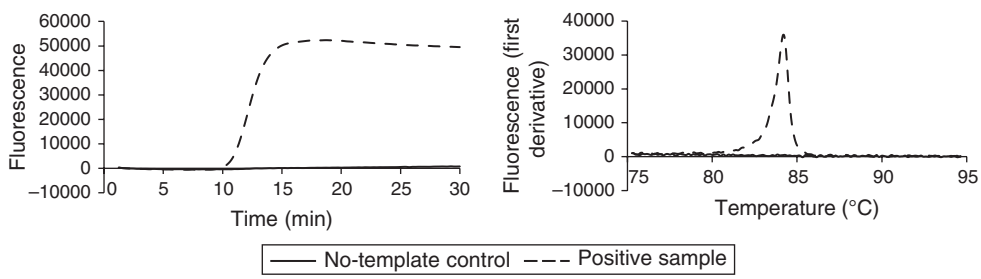


Fig. 4.3. Real-time LAMP results, showing typical amplification plots (left) and annealing plots (right) for positive and negative reactions. (Results for the no-template control are more or less coincident with the baseline.)

achieved by avoiding the variable bases and including degenerate bases in the primers if necessary. LAMP requires the use of internal and external primers, as described above, with the optional addition of loop primers to accelerate amplification. Furthermore, the internal primers (usually referred to as FIP and BIP, the forward and backward internal primers) each consist of two template-binding regions, with the two regions in opposing orientations (Fig. 4.1). Each primer, or template-binding region, is designed according to similar design criteria to those used for PCR primer design, with reference to predicted annealing temperature, lack of secondary structure and potential primer-dimer formation. Annealing temperatures are typically in the range of 55–60°C for F3/B3 (the external primers), the F2/B2 regions in the FIP/BIP primers and the loop primers, and 60–65°C for the F1/B1 regions in the FIP/BIP primers. Target regions are typically in the region of 250–400 base pairs. Specificity is primarily conferred by the internal primers. Due to the opposing orientation of the F1 and F2 regions in the primer FIP (and the B1 and B2 regions in BIP), this means that the most salient bases in terms of specificity are located at the 5′ and 3′ ends of the internal primers. The use of up to six primers targeting eight regions of the target sequence has the potential to confer a high degree of specificity. Primer design software is available (PrimerExplorer from Eiken; LAMP Designer from PREMIER Biosoft) that can be used in the design process.

4.5 Other isothermal methods

While LAMP uses the generation of single-stranded loops to allow primer binding and amplification to occur at a constant temperature, other methods have been developed that use different approaches, some of which allow rapid and efficient amplification of target sequences. Recombinase polymerase amplification (RPA), which uses recombinase and DNA polymerase (Piepenburg *et al.*, 2006), and helicase-dependent amplification (HDA) (Vincent *et al.*, 2004) are amplification methods that use enzymatic approaches to denature dsDNA and allow primer binding. The strand displacement activity of some DNA polymerases is used in methods based on rolling circle amplification (RCA), which involves isothermal amplification of circular templates (Haible *et al.*, 2006). One of the most established isothermal amplification methods is NASBA (nucleic acid sequence-based amplification) (Compton, 1991). This method for amplification of RNA is based on the incorporation of the T7 RNA polymerase promoter sequence into intermediary products using a modified primer. The product of a NASBA reaction is single-stranded RNA, which can be conveniently detected using hybridization, for example, using fluorescent probes such as molecular beacons (oligonucleotide hybridization probes) or array-based methods. NASBA has been used for the detection of a number of plant pathogens in conjunction with molecular beacon probes and real-time fluorescence detection, in a format referred to as AmpliDet (Leone *et al.*, 1998). The mechanism of amplification by NASBA strongly favours the amplification of RNA rather than DNA. As mRNA is less stable than DNA, which degrades rapidly in dead cells, NASBA can be used for the selective detection of viable cells; this can be useful in the context of pathogen detection.

Protocols

Contents

Protocol 4.1

Loop-mediated isothermal amplification (LAMP) with end-point detection

Materials

- LAMP primers. The forward and backward internal primers (FIP and BIP) should be purified by high-performance liquid chromatography (HPLC) (see Note 1)
- *Bst* (*Bacillus stearothermophilus*) DNA polymerase, large fragment, or *Bst* 2.0 DNA polymerase (an *in silico* designed homologue) (both from New England Biolabs)
- $MgSO_4$ (50 mM)
- Betaine (5 M)
- dNTP mix (each at 10 mM)
- Molecular biology grade water
- 1.5 ml Eppendorf tubes (or similar)
- Vortex mixer
- 0.2 ml or 0.6 ml reaction tubes
- Heated block or water bath
- DNA extracted from sample (see Note 2)
- Pipettors and filter tips (different pipettors should be used for setting up reactions, adding DNA to reactions and post-amplification analysis)
- Laminar flow hood for setting up reactions, or a dedicated area for pre-amplification steps

Method

1. Reactions should be set up in a dedicated clean area (ideally in a laminar flow cabinet) to prevent contamination. No DNA and RNA should be handled in this area.

2. The components of a typical LAMP reaction are shown in Protocol 4.1, Table 1.
3. Vortex briefly, then pipette 24 µl aliquots of master mix into 0.2 ml or 0.6 ml tubes.
4. Add 1 µl sample DNA to give a total reaction volume of 25 µl. Do not perform this step in the cabinet used for setting up reactions. Every run should include at least one reaction containing water instead of DNA as a negative control.
5. Incubate the reaction tubes at 65°C for 30–60 min (see Note 4). After this time, incubate the reaction tubes at 80°C for 5–10 mins to inactivate the *Bst* polymerase and stop the reaction.
6. Products can be analysed using a range of end-point methods. Agarose gel electrophoresis can be performed using a 1–2% agarose gel (for a basic protocol, see Sambrook and Russell, 2006).
7. LAMP products should have a characteristic ladder-like appearance when visualized by gel electrophoresis (see Fig. 4.2). Alternative methods for end-point detection of LAMP products include the addition of a fluorescent dye or other indicator, or the observation of turbidity (Tomlinson and Boonham, 2008).

Protocol 4.1, Table 1. Components of a typical LAMP reaction.

Reagent	Starting concentration	Volume per reaction (µl)	Final concentration
Thermopol buffer[a]	10×	2.5	1×
$MgSO_4$[a,b]	50 mM	0–4	2–10 mM
dNTP mix[c]	10 mM each	3.5	1.4 mM
Primer F3[d]	10 µM	0.5	200 nM
Primer B3[d]	10 µM	0.5	200 nM
Primer FIP	100 µM	0.5	2 µM
Primer BIP	100 µM	0.5	2 µM
Primer F-loop[e]	100 µM	0.25	1 µM
Primer B-loop[e]	100 µM	0.25	1 µM
Betaine[b]	5 M	3–7	0.6–1.4 M
Bst polymerase	8 units/µl	1	0.32 units/µl
Water (molecular biology grade)	–	To give a volume of 24 µl	–

[a] 10× Thermopol buffer contains 20 mM $MgSO_4$.
[b] See Note 3 below.
[c] Deoxynucleotide triphosphate.
[d] Forward and backward external primers.
[e] Forward and backward loop primers.

Notes

1. LAMP primers should be designed according to the same principles as PCR primers to avoid secondary structure and dimer formation. The F1 and B1 regions in the FIP/BIP primers should have melting temperatures of 60–65°C, while the F2/B2 regions in the FIP/BIP primers, the F3/B3 external primers and F-loop/B-loop primers should have melting temperatures of 55–60°C. For the discrimination of target and non-target sequences, the most salient mismatches should be located at the 3′ end of F1 and B1 (the 5′ end of primers FIP and BIP). Software for LAMP primer design may also be used (PrimerExplorer from Eiken; LAMP Designer from PREMIER Biosoft).

2. RNA targets can be amplified by reverse transcriptase-LAMP (RT-LAMP) by the addition of a thermostable reverse transcriptase; reverse transcription and polymerization occur at a reaction temperature of approximately 65°C.

3. The concentrations of betaine and $MgSO_4$ can be optimized to maximize assay performance. Betaine concentrations in the range 0.6–1.6 M and $MgSO_4$ concentrations in the range 4–10 mM should be tested. Betaine and $MgSO_4$ concentrations of 1.2 M and 6 mM, respectively, typically give good results with most primer sets.

4. The reaction temperature can be optimized in the range 60–67°C; however, 65°C typically gives good results for most primer sets. A reaction time of 60 min will typically give good results using *Bst* polymerase (New England BioLabs), while a reaction time of 30 min can be used with *Bst* 2.0 (New England BioLabs).

Protocol 4.2

Real-time LAMP

Materials

- LAMP primers (see Protocol 4.1)
- Isothermal master mix (OptiGene)
- Molecular biology grade water
- 1.5 ml Eppendorf tubes (or similar)
- Vortex mixer
- Genie reaction tubes (OptiGene)
- Genie II instrument (OptiGene)
- DNA extracted from sample
- Pipettors and filter tips (different pipettors should be used for setting up reactions and for adding DNA to reactions)
- Laminar flow hood for setting up reactions, or a dedicated area for pre-amplification steps

Method

1. The LAMP reaction mix should be prepared as shown in Protocol 4.2, Table 1.
2. Add 1 µl sample DNA to give final reaction volume of 25 µl. Each run should include at least one reaction containing water instead of DNA as a negative control.
3. Place reaction strips into the Genie II. A typical reaction profile is: amplification for 30 min at 65°C; anneal from 95 to 75°C at 0.05°C/s.
4. Results will be displayed as real-time amplification plots, which can be interpreted in terms of T_p values (the time taken to reach a maximum rate of fluorescence increase) and anneal peaks. A positive reaction is indicated by an exponential

Protocol 4.2, Table 1. Preparation of LAMP reaction mix.

Reagent	Starting concentration	Volume per reaction (µl)	Final concentration
Isothermal master mix	–	15	1×
Primer F3[a]	10 µM	0.5	200 nM
Primer B3[a]	10 µM	0.5	200 nM
Primer FIP[b]	100 µM	0.5	2 µM
Primer BIP[b]	100 µM	0.5	2 µM
Primer F-loop[c]	100 µM	0.25	1 µM
Primer B-loop[c]	100 µM	0.25	1 µM
Water (molecular biology grade)	–	To give a volume of 24 µl	–

[a]Forward and backward external primers.
[b]Forward and backward internal primers.
[c]Forward and backward loop primers.

increase in fluorescence on the amplification plot (a sigmoidal curve, as shown in Fig. 4.3) and an anneal peak at the same temperature as that of the positive control. The anneal temperature is characteristic for each assay, and should fall within approximately ± 0.5°C of the expected value.

References

Arif, M., Ochoa-Corona, F.M., Opit, G.P., Li, Z.H., Kučerová, Z., Stejskal, V. and Yang, Q.Q. (2012) PCR and isothermal-based molecular identification of the stored-product psocid pest *Lepinotus reticulatus* (Psocoptera: Trogiidae). *Journal of Stored Products Research* 49, 184–188.

Boubourakas, I.N., Fukuta, S. and Kyriakopoulou, P.E. (2009) Sensitive and rapid detection of peach latent mosaic viroid by the reverse transcription loop-mediated isothermal amplification. *Journal of Virological Methods* 160, 63–68.

Bühlmann, A., Pothier, J.F., Rezzonico, F., Smits, T.H., Andreou, M., Boonham, N., Duffy, B. and Frey, J.E. (2013a) *Erwinia amylovora* loop-mediated isothermal amplification (LAMP) assay for rapid pathogen detection and on-site diagnosis of fire blight. *Journal of Microbiological Methods* 92, 332–339.

Bühlmann, A., Pothier, J.F., Tomlinson, J.A., Frey, J.E., Boonham, N., Smits, T.H.M. and Duffy, B. (2013b) Genomics-informed design of loop-mediated isothermal amplification for detection of phytopathogenic *Xanthomonas arboricola* pv. *pruni* at the intraspecific level. *Plant Pathology* 62, 475–484.

Compton, J. (1991) Nucleic-acid sequence-based amplification. *Nature* 350, 91–92.

Fukuta, S., Iida, T., Mizumkami, Y., Ishida, A., Ueda, J., Kanbe, M. and Ishimoto, Y. (2003) Detection of Japanese yam mosaic virus by RT-LAMP. *Archives of Virology* 148, 1713–1720.

Gandelman, O., Jackson, R., Kiddle, G. and Tisi, L. (2011) Loop-mediated amplification accelerated by stem primers. *International Journal of Molecular Sciences* 12, 9108–9124.

Gill, P. and Ghaemi, A. (2008) Nucleic acid isothermal amplification technologies – a review. *Nucleosides, Nucleotides and Nucleic* Acids 27, 224–243.

Goto, M., Honda, E., Ogura, A., Notomo, A. and Hanaki, K.I. (2009) Colorimetric detection of loop-mediated isothermal amplification reaction by using hydroxy naphthol blue. *Biotechniques* 46, 167–172.

Haible, D., Kober, S. and Jeske, H. (2006) Rolling circle amplification revolutionizes diagnosis and genomics of geminiviruses. *Journal of Virological Methods* 135, 9–16.

Iwamoto, T., Sonobe, T. and Hayashi, K. (2003) Loop-mediated isothermal amplification for direct detection of *Mycobacterium tuberculosis* complex, *M. avium*, and *M. intracellulare* in sputum samples. *Journal of Clinical Microbiology* 41, 2616–2622.

Kaneko, H., Kawana, T., Fukushima, E. and Suzutani, T. (2007) Tolerance of loop-mediated isothermal amplification to a culture medium and biological substances. *Journal of Biochemical and Biophysical Methods* 70, 499–501.

Kiatpathomchai, W., Jaroenram, W., Arunrut, N., Jitrapakdee, S. and Flegel, T.W. (2008) Shrimp Taura syndrome virus detection by reverse transcription loop-mediated isothermal amplification combined with a lateral flow dipstick. *Journal of Virological Methods* 153, 214–217.

Kikuchi, T., Aikawa, T., Oeda, Y., Karim, N. and Kanzaki, N. (2009) A rapid and precise diagnostic method for detecting the pinewood nematode *Bursaphelenchus xylophilus* by loop-mediated isothermal amplification. *Phytopathology* 99, 1365–1369.

Lenarčič, R., Morisset, D., Pirc, M., Llop, P., Ravnikar, M. and Dreo, T. (2014) Loop-mediated isothermal amplification of specific endoglucanase gene sequence for detection of the bacterial wilt pathogen *Ralstonia solanacearum*. *PLoS ONE* 9(4), e96027.

Leone, G., Van Schijndel, H., Van Gemen, B., Kramer, F.R. and Schoen, C.D. (1998) Molecular beacon probes combined with amplification by NASBA enable homogeneous, real-time detection of RNA. *Nucleic Acids Research* 26, 2150–2155.

Mori, Y., Nagamine, K., Tomita, N. and Notomi, T. (2001) Detection of loop-mediated isothermal amplification reaction by turbidity derived from magnesium pyrophosphate formation. *Biochemical and Biophysical Research Communications* 289, 150–154.

Mori, Y., Kitao, M., Tomita, N. and Notomi, T. (2004) Real-time turbidimetry of LAMP reaction for quantifying template DNA. *Journal of Biochemical and Biophysical Methods* 59, 145–157.

Nagamine, K., Watanabe, K., Ohtsuka, K., Hase, T. and Notomi, T. (2001) Loop-mediated isothermal amplification reaction using a nondenatured template. *Clinical Chemistry* 47, 1742–1743.

Nagamine, K., Hase, T. and Notomi, T. (2002) Accelerated reaction by loop-mediated isothermal amplification using loop primers. *Molecular and Cellular Probes* 16, 223–229.

Nie, X. (2005) Reverse transcription loop-mediated isothermal amplification of DNA for detection of *Potato virus Y*. *Plant Disease* 89, 605–610.

Notomi, T., Okayama, H., Masubuchi, H., Yonekawa, T., Watanabe, K., Amino, N. and Hase, T. (2000) Loop-mediated isothermal amplification of DNA. *Nucleic Acids Research* 28, e63.

Piepenburg, O., Williams, C.H., Stemple, D.L. and Armes, N.A. (2006) DNA detection using recombination proteins. *PLoS Biology* 4, 1115–1121.

Sambrook, J. and Russell, D.W. (2006) Agarose gel electrophoresis. *Cold Spring Harbour Protocols* 2006(1) pdb.prot4020. Available at: http://cshprotocols.cshlp.org/content/2006/1/pdb.prot4020.short (accessed 29 October 2015).

Tani, H., Teramura, T., Adachi, K., Tsuneda, S., Kurata, S., Nakamura, K., Kanagawa, T. and Noda, N. (2007) Technique for quantitative detection of specific DNA sequences using alternately binding quenching probe competitive assay combined with loop-mediated isothermal amplification. *Analytical Chemistry* 79, 5608–5613.

Tomita, N., Mori, Y., Kanda, H. and Notomi, T. (2008) Loop-mediated isothermal amplification (LAMP) of gene sequences and simple visual detection of products. *Nature Protocols* 3, 877–882.

Tomlinson, J. and Boonham, N. (2008) Potential of LAMP for detection of plant pathogens. *CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources* 3, 066.

Tomlinson, J.A., Barker, I. and Boonham, N. (2007) Faster, simpler, more-specific methods for improved molecular detection of *Phytophthora ramorum* in the field. *Applied and Environmental Microbiology* 73, 4040–4047.

Tomlinson, J.A., Dickinson, M.J. and Boonham, N. (2010a) Rapid detection of *Phytophthora ramorum* and *P. kernoviae* by two-minute DNA extraction followed by isothermal amplification and amplicon detection by generic lateral flow device. *Phytopathology* 100, 143–149.

Tomlinson, J.A., Boonham, N. and Dickinson, M. (2010b) Development and evaluation of a one-hour DNA extraction and loop-mediated isothermal amplification assay for rapid detection of phytoplasmas. *Plant Pathology* 59, 465–471.

Tomlinson, J.A., Dickinson, M.J. and Boonham, N. (2010c) Detection of *Botrytis cinerea* by loop-mediated isothermal amplification. *Letters in Applied Microbiology* 51, 650–657.

Tomlinson, J.A., Ostoja-Starzewska, S., Adams, I.P., Miano, D.W., Abidrabo, P., Kinyua, Z., Alicai, T., Dickinson, M.J., Peters, D., Boonham, N. and Smith, J. (2013) Loop-mediated isothermal amplification for rapid detection of the causal agents of cassava brown streak disease. *Journal of Virological Methods* 191, 148–154.

Vincent, M., Xu, Y. and Kong, H.M. (2004) Helicase-dependent isothermal DNA amplification. *EMBO Reports* 5, 795–800.

Wastling, S.L., Picozzi, K., Kakembo, A.S.L. and Welburn, S.C. (2010) LAMP for human African trypanosomiasis: a comparative study of detection formats. *PLoS Neglected Tropical Diseases* 4(11), e865.

Zhang, Z.Y., Liu, X.J., Li, D.W., Yu, J.L. and Han, C.G. (2011) Rapid detection of wheat yellow mosaic virus by reverse transcription loop-mediated isothermal amplification. *Virology Journal* 8, 550.

5 DNA barcoding for identification of plant pathogens

Jennifer Hodgetts*

Fera, Sand Hutton, York, UK

5.1 Introduction

Taxonomy, the classification of organisms, is a profoundly important scientific discipline. Without it, plant pathogen identification or, indeed, precise identification of any organism, would not be possible. Since the advent and uptake of DNA sequencing from the 1980s onwards, analysis of an organism's DNA sequence has become a new tool added to the traditional taxonomic process.

The term DNA 'barcodes' was first used in the scientific literature in the 1990s (Arnot *et al.*, 1993); however, the technique as it is now known came to the forefront with the work of Canadian scientist Dr Paul D.N. Hebert, who is often considered the 'father of DNA barcoding'. In 2003, Hebert and co-workers published a seminal paper 'Biological identifications through DNA barcodes' in the journal *Proceedings of the Royal Society of London B*, which proposed the use of DNA barcoding for species identification (Hebert *et al.*, 2003). Following on from this, DNA barcoding became known as a way to identify species by studying a short standardized gene region, with the principle analogous to the barcode labels used on products in shops. Key to the success of this DNA barcoding approach was the identification of suitable genes, along with standardization of the approach across the different scientific disciplines.

DNA sequencing of loci has commonly been used for much longer for species identification within certain genera or families, although neither the descriptor 'DNA barcode' nor the strict criteria of a DNA barcode were applied. It is also important to distinguish DNA barcoding, an identification tool, from molecular phylogenetics, which is used to determine evolutionary patterns. This delineation enables non-specialists to use DNA barcoding to identify species – one of the benefits of the system that has helped

*E-mail: jennifer.hodgetts@fera.co.uk

to address what has become known as the 'taxonomic impediment'. The taxonomic impediment encompasses the situation arising from lack of trained taxonomists and existing taxonomic knowledge gaps (Goldstein and DeSalle, 2010), which are recognized as barriers to managing biodiversity at all levels. This, along with a very enthusiastic uptake of DNA barcoding by molecular biologists, contributed to substantial early resistance to DNA barcoding from traditional taxonomists, who saw the technique as a threat to taxonomy. This led to many strongly worded publications within the scientific literature, with examples of publication titles including 'The perils of DNA barcoding and the need for integrative taxonomy' (Will *et al.*, 2005) and 'The unholy trinity: taxonomy, species delimitation and DNA barcoding' (DeSalle *et al.*, 2005). Many of these misconceptions and misgivings have been overcome over the past decade, so that numerous close and fruitful collaborations now exist between taxonomists and molecular biologists. Most now believe that this interaction represents the best scenario for the successful application of DNA barcoding, and that the importance of combining traditional morphology with molecular biology cannot be overstated with the two disciplines complementary and working together at their best.

When compared with other diagnostic approaches for a given pathogen, DNA barcoding is unique in that it is a continually evolving discipline, changing on an almost daily basis. This is primarily due to the continual growth of the molecular sequence data sets that act as the backbone of a DNA barcode identification. Researchers are continually generating new sequence data for species previously not studied with molecular tools: for newly identified species, from studying species in more depth, after resolving cryptic species, and from changing and updating taxonomy. Therefore, the application of DNA barcoding to a particular species is never set in stone, as new data may change its usefulness; furthermore, with the publication of new, relevant DNA sequences, each barcode must be regularly evaluated to ensure that the desired outcome is still achievable.

Owing to the very rapid uptake of DNA barcoding, numerous international consortia were established to focus efforts and establish standards, and many are still active. The most widely known is the Consortium for the Barcode of Life (CBOL), which has an overarching focus on the promotion of DNA barcoding. There are also numerous consortia with a specific focus. For example, there are those with a focus on taxa, which include the: Formicidae Barcode of Life Campaign (ants); All Birds Barcoding Initiative (ABBI); Trichoptera Barcode of Life Campaign (caddisflies); Fish Barcode of Life Initiative (FISH-BOL); Lepidoptera Barcode of Life Campaign (butterflies and moths); Mammalia Barcode of Life Campaign (all vertebrates); Mosquito Barcoding Initiative; Marine Barcode of Life campaign (MarBOL); Sponge Barcoding Project (SBP); Bee Barcode of Life Initiative (BeeBOL); Tephritid Barcode Initiative (TBI) (fruit flies); and Shark Barcode of Life project (SharkBOL). Other consortia focus on geographic regions, and include the: Polar Barcode of Life campaign (Arctic and Antarctic ecosystems); Coral Reef Barcode of Life (fish in the Great Barrier Reef); German Barcode of Life (GBOL) network; Norwegian Barcode of Life initiative (NorBOL) and Japanese Barcode of Life Initiative (JBOLI). In addition, other consortia have a specific focus; these

include the Education and Barcode of Life (eBOL), HealthBOL (vectors, pathogens and parasites of human health) and the Quarantine Barcode of Life (QBOL) (European Union (EU) quarantine listed organisms) (Jinbo *et al.*, 2011).

Fundamental to DNA barcoding is the generation of good-quality sequence data linked to a verified voucher specimen to act as reference 'type' sequences. In addition, it is essential to ensure that DNA sequences are publicly available. To this end, a number of databases exist which are repositories for sequence data, the most commonly used being GenBank at the National Centre for Biotechnology Information (NCBI) of the US National Institutes of Health (www.ncbi.nlm.nih.gov/), the European Nucleotide Archive (ENA) at the European Molecular Biology Laboratory (EMBL) (www.ebi.ac.uk/embl/) and the DNA Data Bank of Japan (DDBJ) (www.ddbj.nig.ac.jp/). These are used by scientists worldwide and act to collect annotated sequence data and make them freely available for all. Indeed, the importance of making sequence data accessible is highlighted by journals for scientific publications making this a stipulation. However, the key three databases listed above are not specifically focused on barcoding data and, to that end, BOLD: The Barcode of Life Data Systems (http://www.boldsystems.org/) was created specifically for the storage, analysis and publication of barcodes. While free access to sequence data is vital for the progression of science, there can sometimes be problems associated with this, primarily around the variable quality of published data and the ability to identify errors. Therefore, there are advantages and disadvantages of the main three sequence databases compared with the BOLD database (see Section 5.4, and specifically Section 5.4.3, for further discussion).

Another vital factor for consideration is standardization of the target gene used for DNA barcoding. The key trait required for a barcode is a gene that evolves and mutates at a rate that produces significant changes between different species, but remains relatively conserved within a species. To this end, the mitochondrial gene for cytochrome c oxidase subunit I (*COI*) was proposed for almost all animal groups, with analysis of a 648 bp *5′* fragment. Mitochondrial (and chloroplast) genes represent good targets for PCR-based applications as they are multicopy genes and therefore are easy to amplify using PCR.

During the rapid development and application of DNA barcoding as a specific tool, certain criteria have evolved that encompass and define the requirements of a 'good' DNA barcode. These include both practical and analytical considerations and, in principle, can be summarized as having the following features (Kress and Erickson, 2008; Valentini *et al.*, 2009; Goldstein and DeSalle, 2010):

- standardization of the gene region for taxonomic groups;
- species level variability (at intraspecies and interspecies level);
- justifiable criteria for delimitation of the species;
- a library of reliable sequences (reference sequences, preferably linked to voucher samples);
- conserved flanking sites for universal PCR primers over many genera that provide robust laboratory performance;
- short sequence length (approximately 600 bp) for easy, practical DNA sequencing; and
- standardized methods for comparison of the sequences.

Encompassing these criteria for defining a DNA barcode has led to the creation of data standards that incorporate both specimen information and sequence analysis quality, and allow the definition of a DNA sequence as a 'barcode' in nucleotide sequence databases (Hanner *et al.*, 2009).

A slight variation of traditional DNA barcoding can be the application of principles for the finer differentiation of samples. This generally entails using alternative genes that provide greater genetic variability to study variation within species; for example, to study cryptic species, suspected species complexes or subspecies, or diversity across a geographic range. While this is not strictly DNA barcoding, it can be very useful in certain scenarios, and it can be helpful to combine molecular studies of multiple genes in a process in which sequences from multiple genes are analysed in a combined fashion (concatenated) to provide greater resolving power.

5.2 The application of DNA barcoding

Most diagnostic approaches can be broadly categorized as either specific or generic (non-specific), alternatively referred to as targeted or non-targeted. Typically, the detection strategies used in plant pathogen diagnostics such as ELISA, PCR, real-time PCR and microarrays provide specific detection of a target pathogen (or pathogens). The technique of DNA barcoding straddles both approaches by providing specific identification of the target organism via the use of the generic detection methodology of universal amplification for a group of organisms, e.g. invertebrates, fungi or bacteria. In plant pathogen diagnostics, DNA barcoding is often used to allow specific detection, usually to the species level, occasionally after an initial placement to genus level; therefore, it is a highly flexible and adaptable method which can easily be applied to different diagnostic scenarios.

DNA barcoding is very widely used, and there are examples of the technique being applied across all kingdoms of life. The one major kingdom to which the application of a universal DNA barcode has proved elusive is that of plants, for which the *COI* barcode has proved unsuitable, and finding an equivalent has been challenging. Despite extensive research using numerous targets, no single barcode region with adequate discriminatory power has been identified, and it is now thought that more than one region will be required, as indicated by the CBOL plant working group in establishing two genes (*rbcL* and *matK*) as core barcodes, with supplementary genes where required (Seberg and Petersen, 2009; Hollingsworth *et al.*, 2011).

Even though the application of DNA barcoding is always for either species identification or taxonomic study, the context to which this is applied can vary dramatically; as an example of this diversity, a small subset of these are described here. These examples highlight the use of barcoding in identifying partial or damaged specimens and processed material, where traditional taxonomic study may not be possible. DNA barcoding has been used to identify cases of misidentification of food products in fish to identify substitutions which may either have health implications for consumers or have been made fraudulently for financial gain (Yancy *et al.*, 2008; Barbuto *et al.*, 2010). Siddall *et al.* (2012) studied nematodes as parasites from kosher-certified fish

products to determine whether or not the required food preparation standards had been followed, an application of barcoding to a cultural concern. Barcoding is often employed to aid biodiversity studies, where it often indicates higher than expected levels of biodiversity, and is frequently linked to conservation efforts (Francis *et al.*, 2010). It has also been used to study trophic links at all levels (Hrcek *et al.*, 2011), and applied to museum collections to provide access to historical records (Wandeler *et al.*, 2007; Puillandre *et al.*, 2012). Holmes *et al.* (2009) applied the method to shark and ray fins collected from vessels that had been caught illegally fishing and proposed its use by fishing authorities to assist in management and conservation. In biosecurity, barcoding has been proposed by the EU for the identification of listed plant pests and pathogens (van de Vossenberg *et al.*, 2013) and has been studied in New Zealand as a tool to help in the biosecurity of the country against alien invasive species (Armstrong and Ball, 2005). Indeed, Manghisi *et al.* (2010) demonstrated the use of DNA barcoding to allow easy identification of the introduction of alien species of macroalgae into Sicily.

5.3 Technical considerations

Conducting DNA barcoding is a simple process both conceptually and practically. DNA is extracted from the sample, PCR is performed using primers for a gene appropriate for the type of sample, the PCR product is visualized by agarose gel electrophoresis and then purified. The PCR amplicon is then subjected to DNA sequencing. It should be noted that there are many important factors which require careful consideration for the technique to be applied successfully, and permutations that may vary depending upon the specific application in hand – either the creation of reference sequences to develop a barcoding protocol or, alternatively, the application of barcoding to produce an identification for a sample using an established protocol. The key areas for consideration are discussed below in Sections 5.3.1–5.3.8.

5.3.1 Sample collection and identification

The most crucial aspect of any DNA barcoding experiment commences before the molecular biology laboratory with, first, the collection of the sample, and second, the identification of the sample. Sample collection should be done with care to determine as much information on the origin and nature of the sample as possible including, in the case of plant pathogens, the identity of the host, the country and geographic region of origin, and the symptoms displayed by the host. If the sample is to be used as a voucher specimen, then identification by the relevant taxonomic methods is required. However, in the application of DNA barcoding for sample identification, this is not essential, although it is desirable, as it may provide additional information to add robustness to the final identification.

5.3.2 DNA extraction

Key to the success of DNA barcoding is the extraction of sample DNA of an acceptable standard. A wide range of DNA extraction methodologies exist, and many are suitable for DNA barcoding applications, with the choice often dependent upon the nature of

the sample. Generally, DNA extractions can be broadly categorized as crude or refined. Crude DNA extractions are usually very rapid, including simply boiling tissue in buffers, or chelex resin-based protocols. These typically provide very crude DNA extracts with lower DNA yields which, while suitable for PCR, often cannot be stored over a long period. As a result, these methods are undesirable for voucher specimens, but are acceptable to enable the identification of samples. Alternatively, refined DNA extractions such as spin column- or magnetic bead-based commercial kits or the laboratory 'gold-standard'-CTAB (cetyltrimethyl ammonium bromide) buffer-based protocols produce much purer, high-quality DNA which can be stored over the long term. Therefore, these approaches are preferable for voucher samples which may be required for further analysis in the future. Consideration should also be given to the size of the sample (see Section 5.3.3) and optimization or scaling of the chosen protocol so that it is appropriate for the sample size. This generally requires reduction in the elution volume for the DNA extract.

5.3.3 The sample and sample matrix

One element that is highly variable in DNA barcoding is the nature of the sample and, in addition, the sample matrix. For example, DNA barcoding can be used for invertebrates (insects and nematodes, or parts thereof), vertebrates, plant material (leaves, stems or roots of plants or produce) or bacterial or fungal cultures, to name but a few. A primary method of differentiation for DNA barcoding samples is whether the sample for identification is in a pure form or within a matrix, e.g. a bacterial culture (a pure sample) as opposed to the bacterium within its host tissue (a sample within a matrix). This is an important consideration. If the test subject is within a host background, then the laboratory methods used may have to be modified. In certain circumstances, it may be possible to remove the matrix, for instance a bacterial or fungal pathogen may be cultured from the host tissue, and the pure culture subject to DNA barcoding. If this is possible, then it is recommended, as a pure sample simplifies the DNA barcoding process. There are, though, instances when testing is always conducted within a background sample, such as the case of non-culturable pathogens, including phytoplasmas. When DNA barcoding a sample within a host matrix, it is important that the DNA extraction method selected is optimized to extract, as a minimum, both the host and the target tissue, although preferably the method should preferentially target the sample rather than the host. For instance when testing for a bacterium within an insect, a method should be selected that is optimized for the extraction of bacterial DNA.

Another factor to consider is the sample size. PCR is a very sensitive method, so the amount of starting material, given a suitable DNA extraction method, can be very small. In the case of invertebrate samples, only a single leg or body part is required. Indeed, barcoding is often applied to invertebrate samples that are damaged, or when only a partial sample is available, in which case traditional taxonomic identification is not possible. Due to the nature of the DNA sequencing method used when barcoding, it is most important that each sample is composed of only organism, e.g. a culture is pure and not composed of multiple species. Therefore, taking a small sample can help to avoid mixed samples.

The final major consideration for the sample is the number of samples tested, paying particular attention to variation with the species. This is especially important for voucher samples, but is also pertinent for samples for identification. It is important to consider that certain species may be present in complexes, or in mixed infections where there may be a dominant species. As in plant pathogen detection, it is important to detect all pathogens present, so a good depth of sampling is required. In addition, it is important to test several samples from the same source to ensure that variation within the population does not bias the identification or cause blurring between species boundaries.

5.3.4 Primer selection

Primer selection is vitally important to the success of DNA barcoding and is often the most challenging aspect. Invertebrate DNA barcoding uses a fixed gene as the initial barcoding marker, while the DNA barcoding of bacteria, fungi and phytoplasma frequently uses suites of primers to achieve a tiered identification. Thus, given a sample for which no prior identification has been made, for example it is only known that it is a fungal sample, it is likely that, first, a gene will be used for generic placement, and then another gene will be used that is suitable for differentiation within the given family.

Once the required target marker has been established, then the best PCR primers for the given sample must be identified. Typically, for DNA barcoding, generic primers that can amplify a broad range of species or genera are used. However, no single primer pair will amplify all species, so often a small number of primer pairs are used which, between them, will provide amplification of the majority of species. To enable amplification of a wide range of species, primers are first designed to conserved gene regions, and secondly, primers often contain degenerate nucleotide positions, they enable the priming of variable sites.

As an alternative to screening samples with numerous primer pairs, primer pools can be used; in this approach, multiple primers for the same target region are combined and used within a single PCR reaction. The approach can be useful in other diagnostic techniques, but it is generally not recommended for DNA barcoding, simply because DNA sequencing reactions need to be performed with the same primers as the PCR, and if a pool of primers was used, it would not be possible to determine which primer initiated amplification.

A final consideration for PCR primer selection is whether the sample is being tested within a background of host or contaminating DNA. When testing for sample in a background, it is essential to ensure that the selected PCR primers will not interact or amplify DNA from the matrix either in addition to or instead of the target sample. Depending upon the situation and the gene, this may require the selection of different primers to target the sample more specifically.

5.3.5 Polymerase selection

According to the nature of the barcode region being studied, identification to the species level can be based on a very small number of nucleotide positions, so accurate amplification during PCR is critical to success.

To allow accurate amplification, polymerases commonly referred to as proofreading or high-fidelity polymerases are used. These have 3′ to 5′ proofreading exonuclease activity whereby, during amplification, each base is checked and any incorrectly incorporated bases excised and the correct base reincorporated. This provides a very high fidelity to the amplification process, reducing the incorporation of errors that may bias later identification. The use of non-proofreading polymerases without this capability should be avoided for DNA barcoding, as errors introduced during PCR can instead be attributed to the sequence of the sample. This can have a range of negative effects, ranging from misidentification to erroneously high levels of variation, or deceptively suggesting the presence of a new species. While proofreading polymerases are now widely available, the main drawback to their use is the typical requirement for greater optimization of PCR reactions to ensure good performance. Nevertheless, this should be undertaken because of the risks posed by sequence errors in DNA barcoding-based identification.

5.3.6 Polymerase chain reaction (PCR)

For DNA barcoding, it is typical for the PCR cycling conditions to have a lower annealing temperature than would normally be expected in order to enable the generic amplification of a broad range of species. This is especially the case when testing samples that are not in a background of host DNA, when the risk of primer cross-reactivity is reduced. In addition, it is common for a high number of cycles to be used compared with other PCR applications. Both of these factors can lead to an increase in non-specific amplification, so it is important that assays are carefully optimized. Higher DNA concentrations enable the use of fewer PCR cycles, so DNA extraction protocols that generate purer, high-quality DNA extracts can be favourable to ensuring better PCR performance. Occasionally, adjuvants such as BSA (bovine serum albumin), DMSO (dimethyl sulfoxide) or glycerol are added to the PCR; these act to improve the assay performance, particularly with challenging template DNA.

Due to the nature of Sanger sequencing (see Section 5.3.8), only a single amplicon can be sequenced in a reaction, so assays that produce non-specific amplification are not desirable as this necessitates the use of gel extraction purification to remove the non-specific amplicon prior to DNA sequencing. Although this is a simple process which produces good results, it is relatively labour intensive, and optimization of the assay can be used to remove the necessity for this step.

Occasionally, certain sample types require the use of nested PCR. This entails the use of two sequential rounds of PCR, where the product from the first is used as the template for the second. This approach is used when the levels of target DNA are inadequate to allow the production of PCR amplicons in a single round of PCR. The majority of samples for DNA barcoding do not require the use of nested PCR, with the exception of phytoplasmas, for which its use is standard due to the very low titre of the pathogen within the hosts. See Chapter 2 for further information on PCR.

5.3.7 Whole genome amplification

Whole genome amplification (WGA) encompasses a range of techniques which aim to increase the amount of DNA from a small, finite sample by replicating it in a non-target

specific fashion, producing DNA as close as possible to the starting material. This is deployed in situations when there is either a limited amount of DNA or an amount insufficient for the given application. The methodologies used for WGA fall into two main categories, each suited to a different application. The first encompasses approaches that utilize degenerate or random primers; these are variations on PCR that amplify DNA in a non-specific way. Alternatively, multiple displacement amplification (MDA) is an isothermal process that uses random hexamers and a specific type of polymerase (Φ29, from *Bacillus subtilis phage* phi29) that has strand displacement activity.

To date, none of the techniques available are able to create identical additional DNA from a sample; the copied DNA contains minor introduced sequence errors, with loss of coverage of some genes and sequence regions. Therefore, these techniques should only be applied to DNA barcoding samples as a last resort, and when they are applied, it is vital to ensure that numerous populations of a species are tested to ensure that introduced errors are not attributed to a species barcode.

5.3.8 DNA sequencing

For DNA barcoding, chain termination dideoxynucleotide sequencing, commonly known as 'Sanger sequencing', is the primary technique used. Indeed, as DNA barcoding relies inherently on Sanger sequencing, the amplicon length of standardized DNA barcodes is generally within the 600–800 bp read length that can be achieved in a single sequencing run.

When the appropriate equipment is available, DNA sequencing can be performed in the molecular biology laboratory, and the protocols are simple to conduct. None the less, due to economies of scale, it can often be considerably cheaper for the sequencing to be conducted by commercial companies. This also provides additional benefits, such as quality control, which may be desirable if the DNA barcoding is being used for identification purposes. It is important when both producing an identification by barcoding or creating reference sequences that at least two sequence reads are generated for each amplicon. This is usually achieved by the amplicon being sequenced on each strand using both PCR primers, thus ensuring that an accurate sequence is generated for analysis.

An important consideration for the success of DNA sequencing is primer selection. Most often, for the standard *COI* barcode, the PCR primers are used for the sequencing reaction. Occasionally, PCR primers may have inherent characteristics such as secondary structure, which, while adequate for PCR, lead to poor performance during sequencing. In these instances, specific sequencing primers are used that are internal to the PCR primers, but which have characteristics specifically designed to enable high-quality sequencing.

5.4 Analytical considerations

An area of DNA barcoding which is often overlooked in comparison to the laboratory experiments is the analysis of sequence data. However, both the preparation of the sequence data for analysis and the analysis itself are important to the reliability of

a DNA barcode-based identification. The process is multifaceted, encompassing both preparation and analysis of the produced sequence data, and the evaluation of the applicability of the barcode region for the target species. If these aspects are not conducted with careful consideration, then DNA barcoding can produce misidentifications.

5.4.1 Analysis of sequence data

Once the DNA sequencing reaction has been completed, a DNA sequence read, sometimes called an electropherogram, chromotogram or trace file, is generated that requires diligent analysis (see Fig. 5.1). To analyse a sequence read, specialized software is required. Some programs are simple, allowing viewing and basic editing of the trace files, and a number of these are freely available. Commercially available software packages that can be purchased are often much more advanced, providing features such as assembly of reads, alignments and phylogenetic analysis. In essence, the user views the raw sequence data and ensures that the base call generated by the software for each nucleotide position is correct, paying particular attention to features that can be problematic for software programs to analyse. These include features such as homopolymer runs (long runs of a single nucleotide), polymorphic bases (single bases positions where a mix of nucleotides are present) and determining the start and end of a 'good' quality sequence. This latter point is important, as owing to the inherent nature of Sanger sequencing, the first 30–50 bases of the sequence read are not adequately resolved and cannot be used. Typically, a single read generates 600–800 base pairs of sequence, the quality of which will decline at its extremities. Determining the point at which each base call is of adequate quality is important. As a rule, software programs use Phred quality scores; these are assigned to each base and indicate the probability of an incorrect base call; they are often displayed pictorially. A Phred score of 20 is typically the cut-off for a high-quality sequence, giving 95% accuracy. The higher the Phred score, the higher the accuracy for the given base, i.e. a Phred score of 40 has 99.99% accuracy. This information can be used as a guide to trim a sequence read (especially at the 3′ end where the quality can be low) to retain only high-quality, reliable sequence data. Once sequence reads have been checked and trimmed by the user, the sequence data are ready for further analysis.

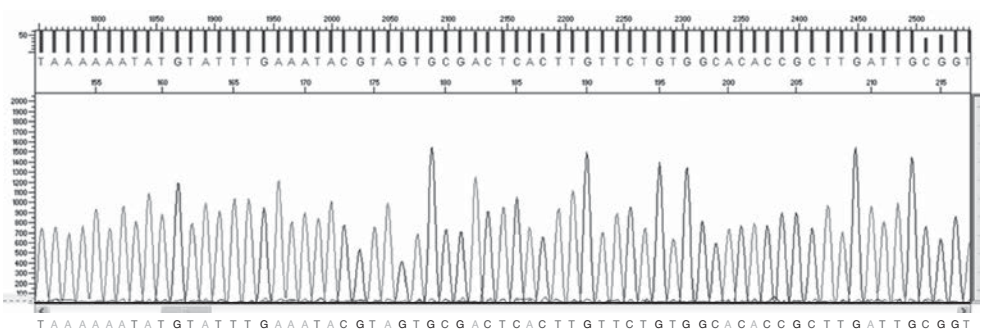


Fig. 5.1. An example of a DNA sequence read, visualized in the software program Sequence Scanner (Applied Biosystems). Note the vertical bars at the top are a diagrammatic indication of the Phred quality score.

5.4.2 Consensus sequences

In DNA barcoding, it is imperative that the precise DNA sequence of the target region is generated. For this reason, it is standard practice for the DNA sequencing of the amplicon to be performed with both PCR primers so that the entire amplicon is sequenced twice, described as achieving two times coverage, with the resultant sequence described as the 'consensus' sequence. During creation of the consensus sequence, the reads generated from both primers are compared so that any ambiguities can be identified and resolved, thus ensuring 100% confidence in the consensus sequence. There is some debate as to exactly what the consensus sequence should include. In an ideal scenario, it would include two times coverage of the entire amplicon from both primer sites; but in practice, due to the loss of the initial bases of the sequence, this is not achieved unless additional processes such as cloning of the amplicon and sequencing of the insert within a vector are undertaken. Therefore, in practice, many users rely on the fact that the beginning of sequence reads are typically very high quality to create a consensus where the majority of the sequence contains two times coverage; however, the start and end have only single read coverage (see Fig. 5.2). Once a consensus sequence has been created, it is important to remove the sequences of the PCR primers which are regions of sequence incorporated or copied from the primers rather than being sequenced from the sample itself.

Degenerate bases arise when the precise nucleotide at a given position cannot be accurately determined from the sequence read data. Instead, a degenerate nucleotide determined from a standardized code encompassing the possible variants is used in its place. In DNA barcoding, due to the importance of accurate sequence, it is preferable not to use degenerate bases, and often the base can be determined by repeated sequencing of the amplicon, or by comparison of the two reads generated; nevertheless, there are cases when there are genuine polymorphisms in the target DNA, especially with multicopy genes.

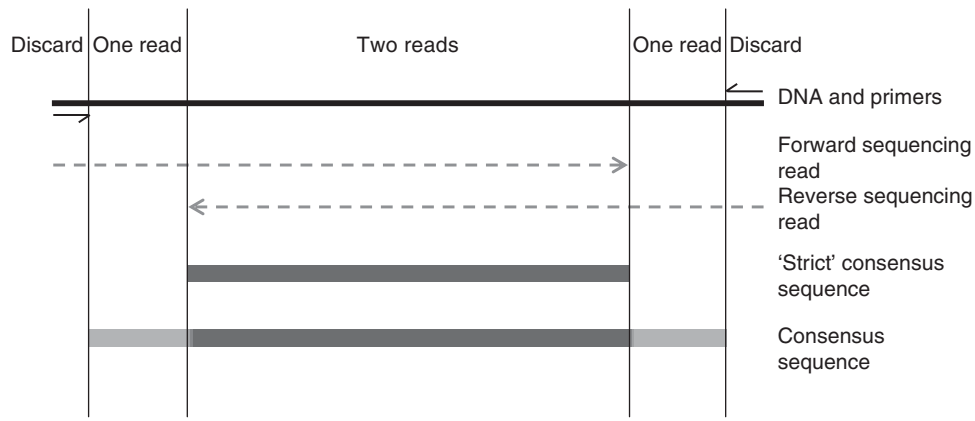


Fig. 5.2. Illustration of the different types of consensus sequences. In the consensus sequences, dark grey indicates regions where two times coverage is achieved (generated from a compilation of sequence reads generated from both primers) and light grey indicates regions with one times coverage. Note the primer binding sites are removed from the consensus sequences in all scenarios.

5.4.3 Analysing a DNA barcode

Once a consensus sequence has been generated, the process that embodies DNA barcoding for identification is undertaken. Primarily, the identification of a species from a barcode can be undertaken via two interlinked processes; either by using a database search or by creating a phylogenetic tree. Sequence similarity searches against databases are the primary tool used for identification of DNA barcodes, so an awareness of the reference sequences being used for this search is important in order that the validity of the search results can be assessed (see Section 5.4.4 for further discussion).

Before using publicly available data sets for identification, it is important to understand the nature of each of these databases and the benefits and drawbacks of each. In essence, the tools used to analyse a DNA barcode fall into two categories: sequence databases such as NCBI, EMBL and DDBJ, as opposed to BOLD and Q-Bank (www.q-bank.eu; a European initiative which is a collection of comprehensive databases on quarantine plant pests and diseases). The key, but vital, difference between these two categories of database lies in the curation of the data; BOLD and Q-Bank only contain sequences curated to minimum quality criteria and all of the sequences are linked to voucher samples/specimens. This process of curation acts to remove a large number of errors which can be present in other data sets. Indeed, the primary drawback of databases such as NCBI, EMBL and DDBJ, in which there is no requirement for sequences to be linked to voucher samples, is the issue of misidentified or incorrect sequences. There is a wide range of reasons that published sequences can be erroneous; misidentification of the sample prior to sequencing is a major contributor, and the curation in BOLD and Q-Bank acts to remove this as a source of errors. Nevertheless, laboratory errors such as mislabelling or the mixing up of samples, the use of poor-quality sequence data and contamination of the sample can also lead to inaccurate sequence data.

With its particular focus of DNA barcodes, BOLD is a useful resource and provides additional valuable features, such as a primer database. The specific aim of BOLD, unlike the other main databases, is to combine in one place the molecular, morphological and distribution data for a specimen, thus assisting in the collation of information that meets the requirements of a barcode designation. However, it is somewhat limited in use for plant pathogen diagnostics, as the focus is on identification of animal, plant and fungal specimens (using the loci of *COI*, *matK* and *rbcL* and the internal transcribed spacer (ITS) region, respectively). Nevertheless, the ethos and standards embedded in BOLD are aspirational for the generation and publication of high-quality DNA barcodes. These principles have been taken forward in an application that is of specific use for plant pathogen identification, with the creation of Q-Bank, which provides curated DNA barcodes of EU quarantine-listed plant pests and pathogens, utilizing similar criteria to BOLD (Bonants *et al.*, 2013).

The generation of phylogenetic trees can also be highly informative, and sequences can be downloaded from NCBI, EMBL and DDBJ, so that a user can create trees using relevant species. While BOLD and Q-Bank both create phylogenetic trees from the unknown sample and the reference database, the sequences cannot be downloaded to enable the end users to conduct their own phylogenetic analyses, a minor drawback of these databases.

5.4.4 Critical factors for evaluation of a DNA barcode

Fundamental to the use of barcoding as an identification tool is the presence of what is often referred to as 'the barcode gap', which ensures that interspecific variation (between species) exceeds intraspecific variation (within species), thereby allowing unambiguous species-level identification (Weimers and Fiedler, 2007). The barcode gap can be determined by percentage sequence similarity scores, and is often visualized using phylogenetic trees. A major factor that can affect the determination of the presence or absence of a barcode gap is the amount of relevant data available, and a major criticism of barcoding as a stand-alone method for identification is that a barcode gap can be simply an artefact of inadequate sampling (Weimers and Fiedler, 2007). What may appear as a barcode gap could be a reflection that for a given group sequence data are not available for sufficient species within the genus, or that there is inadequate representation of the species to capture the natural variation within it. For well-studied organisms, this is rarely a problem, but for others it can prevent the use of DNA barcoding for identification. Although it is difficult to assess whether enough data are generally available, it is desirable to have sequences from multiple populations of the species concerned from across its natural range and to avoid the risks posed by misidentifications; this can be done by using data generated from multiple laboratories (where possible), which add robustness to the data set.

The presence of the barcode gap for a given analysis is also an aspect of barcoding that requires regular evaluation. As new sequences are generated and published, this can affect the outcome of an identification either positively or negatively, that is, new sequences can either remove or create a barcode gap. This is something that must be evaluated regularly, and it sets barcoding apart from other diagnostic methodologies in which assay performances may be not be reviewed regularly.

5.5 Barcoding plant pathogens

The groups described in the following sections each contain numerous important plant pathogens, many of which fall under EU regulation. Each group is very large and diverse, with both the taxonomy and the nature of the pathogens and the diseases they cause under continuous study. Therefore, the diagnostic protocols required for precise identification can evolve over time, as new, more informative markers are identified. To allow DNA barcoding of these groups, careful analysis of the current state-of-the-art is required at regular intervals and, as a result, fixed protocols are not provided.

5.5.1 Bacteria

Barcoding of bacteria is often complicated by the taxonomic classification frequently going beyond the species level to subspecies or pathovar designations, for example in the genera *Xylella*, *Xanthomonas* and *Pseudomonas*, which all contain species that are important regulated plant pathogens. Hence, for regulated organisms, identification to the pathovar level may be necessary. Due to this diversity, the target gene region often varies depending on the genus being studied. The traditional marker used for bacterial identification, the 16S rRNA gene, is typically used to allow a broad

classification to the genus level. This may then require the sequencing of alternative genes (sequentially if necessary) for species-specific identification, with the particular combination of genes required dependent upon the study. Common genes used for bacteria include the *rpoD*, *gyrB*, *DnaA* and *tuf* genes.

5.5.2 Phytoplasmas

As phytoplasmas are non-culturable organisms, their taxonomy is based on sequence analysis of their 16S rRNA and, in fact, may be considered an application of DNA barcoding. However, a 1250 bp region of the gene is required for taxonomic placement to the species level, and this presents drawbacks for the use of this marker as a classical barcode, as it requires multiple sequence reads to generate the full-length sequence. Because of this, recent studies have proposed shorter regions such as the elongation factor EF-Tu (*tuf*) and *secA* genes as alternate barcodes for phytoplasmas, which both provide species-level discrimination (Hodgetts *et al.*, 2008; Makarova *et al.*, 2012). Numerous other genes have been studied for finer differentiation of phytoplasma strains, including the ribosomal protein (rp) operon genes, *secY*, *imp* and *amp* genes. These genes are often highly informative, however the major drawback is the lack of universal PCR primers, which prevents their use as a DNA barcode (Foissac *et al.*, 2013) (reviewed in Hodgetts and Dickinson, 2010).

5.5.3 Fungi

Attempts to employ the standard *COI* barcode to fungi have been of limited success; the presence of introns complicates PCR, designing universal PCR primers is difficult, the gene can be highly variable in size, so complicating analysis, and it can lack adequate sequence variation. Therefore, the ITS region, a nuclear gene, is commonly used as a standard fungal barcode region, aided by being present in multiple copies in the genome and allowing good taxonomic differentiation. The elongation factor 1α (EF1, or TEF1) gene is also commonly used, and again for certain species alternative genes may be required for finer differentiation (see reviews in Seifert, 2009, and Schoch *et al.*, 2012).

5.5.4 European regulated plant pathogens

Over recent years, a concerted effort has been under way to develop DNA barcoding protocols for EU-regulated plant pests and pathogens, under the EU-funded QBOL project, leading to the development of Q-Bank (see Section 5.4.3) (Bonants *et al.*, 2010, 2013). Unfortunately, no single barcode region is suitable for the identification of all regulated organisms and, indeed, many organism groups require the sequential or parallel study of two or more gene regions (van de Vossenberg *et al.*, 2013). However, as the technique shows great promise, barcoding protocols have been harmonized and a draft EPPO (European and Mediterranean Plant Protection Organization) diagnostic standard is under preparation to allow the technique to be readily applied to these organisms.

5.5.5 Establishing a new barcoding protocol

There may be instances in which a DNA barcoding protocol is desired, yet there are no reference sequences available. In this scenario, it may be possible to create reference sequences to determine whether DNA barcoding would be suitable. Essentially, this requires the acquisition of taxonomically identified samples of genera and species closely related to the target; various markers of these closely related taxa are sequenced to act as reference sequences. Fundamentally, the difference between intraspecies and interspecies variation is evaluated (see Section 5.1). It is vital that both an adequate depth and geographic coverage of sampling are achieved in order to ensure that intraspecific variation is considered in the evaluation of whether or not the marker is suitable for species identification (Bergsten *et al.*, 2012). Practically speaking, the amplicon should be around 600 bp in length and provide robust performance under PCR and sequencing. It is desirable to comply with existing standard markers for an organism group, to allow ease of use by the wider scientific community and, if new applications or markers are identified, to publish the methods and reference sequences in the scientific literature.

Protocols

Contents

Protocol 5.1

Polymerase chain reaction and DNA sequencing

Materials

- Proofreading DNA polymerase and PCR buffer (usually supplied with the polymerase), molecular biology grade water. $MgCl_2$ and dNTPs (deoxynucleotide) if polymerase is not in a master-mix format
- Target specific primers – forward, reverse and sequencing (if applicable)
- Molecular grade agarose, TBE (Tris-borate-EDTA) buffer and intercalating dye (such as ethidium bromide, SYBR Green or GelRed)
- PCR purification kit or gel extraction kit (spin-column based)
- Pipettes, three sets (one each for setting up PCR reactions, adding DNA to PCR reactions and post-PCR analysis)
- Filter tips for pipettes
- Vortexer and microcentrifuge
- UV PCR cabinet for setting up PCR reactions
- Thermocycler
- Equipment for agarose gel electrophoresis
- Molecular size/weight marker (for agarose gel)
- Access to Sanger sequencing provision

Method

1. Establish the likely required gene for species identification of the test sample (as determined from the current literature/diagnostic protocols). For a sample with no *a priori* knowledge as to genus (i.e. only the pathogen type is known), the following genes are typically used for initial assessment:

(i) for placement to genus in bacteria, a 350 bp section of the 16S rRNA gene;

(ii) for placement to species in phytoplasmas (in the majority of cases), the 16S rRNA gene (a 1250 bp section);

(iii) for placement to species in fungi in the majority of cases (or occasionally to genus), the ITS region (including the ITS1, ITS2 and 5.8S rRNA).

2. Extract DNA from the sample using an appropriate method (see Section 5.3.2). In the majority of cases, this would entail a CTAB (cetyltrimethyl ammonium bromide)- or spin-column-based method, depending on the nature of the sample.

3. Perform PCR following standard protocols for the given gene (see Protocol 5.1, Table 1). The primary variables which may affect the success of the PCR, and which can be optimized to improve performance if required, are the quantity of template DNA added per reaction, the primer concentrations used, the primers selected, the annealing temperature and the elongation time:

Protocol 5.1, Table 1. Barcoding primers for bacteria, phytoplasmas and fungi for initial taxonomic placement.

Primer name	Sequence (5′–3′)[a]	Amplicon size	Reference
Bacteria 16S rRNA gene			
pA	AGAGTTTGATCCT GGCTCAG	1500 bp (amplified by pA/pH, of which 350bp is sequenced using primers 358-339 and 536-519)	Edwards *et al.*, 1989 Coenye *et al.*, 1999
pH	AAGGAGGTGAT CCAGCCGCA		
358-339	ACTGCTGCCTCC CGTAGGAG		
536-519	GTATTACCGCG GCTGCTG		
Phytoplasma 16S rRNA gene			
P1	AAGAGTTTGATCC TGGCTCAGGATT	1250 bp (amplified by nested PCR using primers P1/P7 in the first round PCR producing a 1800 bp amplicon and R16F2n/R16R2 in the nested round PCR)	Deng and Hiruki, 1991 Smart *et al.*, 1996 Gundersen and Lee, 1996
P7	CGTCCTTCATCGG CTCTT		
R16F2n	GAAACGACTGCT AAGACTGG		
R16R2	TGACGGGCGGT GTGTACAAACCCCG		

Continued

Protocol 5.1, Table 1. Continued.

Primer name	Sequence (5′–3′)[a]	Amplicon size	Reference
Fungi ITS region (including the ITS1, ITS2 and 5.8S rRNA)[b]			
ITS4	TCCTCCGCTTA TTGATATGC	550 bp	White *et al.*, 1990
ITS5	GGAAGTAAAAG TCGTAACAAGG		

[a] A, adenine; T, thymine; G, guanine; C, cytosine.
[b] Only pure cultures should be used with this primer pair (not infected plant material).

(i) PCR reactions should be prepared as described in Protocol 5.1, Table 2 in a UV PCR cabinet in a DNA-free area with dedicated pipettes;

(ii) reagents should be thawed and vortexed to mix prior to preparing a master mix for the number of samples and a small excess;

(iii) once prepared, vortex the master mix to mix and dispense 24 µl aliquots into a 96 well plate or PCR strip;

(iv) remove the reactions to a further area for the DNA to be added to the reaction (with another set of dedicated pipettes). A no-template control should be prepared (with molecular grade water in place of DNA) and a positive control for the given set of primers;

(v) place the PCR reactions into a thermocyclyer (with heated lid). PCR cycling conditions are as follows:

(a) for the bacterial 16S rRNA gene, 4 min at 95°C, 3 cycles of 45 s at 95°C, 2 min at 55°C, 1 min at 72°C, 30 cycles of 20 s at 95°C, 1 min at 55°C, 1 min at 72°C, final extension for 7 min at 72°C;

(b) for the phytoplasma 16S rRNA gene first round PCR, 3 min at 95°C, 35 cycles of 30 s at 94°C, 90 s at 53°C, 90 s at 72°C, final extension for 10 min at 72°C; for the nested round, 3 min at 95°C, 35 cycles of 30 s at 94°C, 90 s at 56°C, 90 s at 72°C, final extension for 10 min at 72°C;

(c) for the fungal ITS gene, 5 min at 94°C, 40 cycles of 45 s at 94°C, 30 s at 52°C, 90 s at 72°C, final extension for 6 min at 72°C.

4. In a further dedicated post-PCR area, perform agarose gel electrophoresis using standard protocols to gauge the success of the PCR. Gel percentages are typically 0.8–2.5%, with increased percentage as product size decreases. A molecular marker of an appropriate size range should be used for each PCR amplicon. Ensure that only the specific (expected) amplicon is produced and gauge the concentration of the PCR amplicon by comparison with the marker.

Protocol 5.1, Table 2. The components of a PCR reaction.

Component	Working solution	Final concentration	Volume per reaction
Proofreading DNA polymerase master mix[a]	2×	1×	12.5 µl
Molecular grade water	n/a	n/a	10.5 µl
Forward primer[b]	10 µM	0.2 µM	0.5 µl
Reverse primer[b]	10 µM	0.2 µM	0.5 µl
DNA[c]	n/a	n/a	1 µl

[a] Various proofreading DNA polymerases can be used. These can be purchased as a pre-mixed master-mix format or as individual components requiring the separate addition of buffer, $MgCl_2$ and dNTPs. Master mixes offer increased convenience and reduce the risk of pipetting errors and contamination. Different DNA polymerases may give better performance for given assays.
[b] The final primer concentration may vary depending on the assay, usually ranging from 0.1 to 0.6 µM.
[c] In the majority of cases, the DNA can be used with the concentration as extracted and does not require quantification or adjustment.

5. Purify the PCR amplicon using a PCR purification kit (spin-column based) following the manufacturer's protocol, e.g. Qiagen's QIAquick PCR purification kit. The final elution volume may require adjusting so that the purified PCR amplicon is in the required concentration range for the sequencing provider.
6. If the PCR produced multiple amplicons (the specific desired amplicon and non-specific amplicon(s)) then perform gel extraction and purification of the specific amplicon following the manufacturer's protocol, e.g. using Qiagen's QIAquick gel extraction kit.
7. Submit the purified amplicon and a small amount of each PCR primer for DNA sequencing to a DNA sequencing provider following the provider's guidelines. Each sample should be sequenced on both strands.

Protocol 5.2

Sequence analysis and identification

Materials

- Sequence analysis software for viewing, editing and trimming Sanger DNA sequence reads and creating consensus sequences from multiple reads
- Access to nucleotide sequence databases and/or phylogenetic analysis software

Method

1. Retrieve DNA sequence reads from sequencing provider.
2. Using the available software, examine the sequence read to confirm the base calls (nucleotides), as below:
   (i) If the software allows, align the forward and reverse sequence read files and check for any anomalies in each read and in regions where two reads are present. Correct if necessary, remove primer sites if present and export the checked and edited consensus sequence.
   (ii) If the software available does not allow alignment of multiple sequence reads, check and correct the sequence read from each primer individually, and export the checked and edited sequences generated from each primer. Align the two sequences and check for any anomalies between the two reads in the overlapping regions (if any are found, resolve these by checking back to the sequence reads), and then create the consensus sequence by combining the two reads into one file.
3. Export the checked and corrected consensus sequence from the software (typically in FASTA format) for analysis.
4. Perform identification of the sequence by:
   (i) searching the sequence in the relevant database(s); and/or
   (ii) performing phylogenetic analysis with a relevant set of reference sequences.

References

Arnot, D.E., Roper, C. and Bayoumi, R.A.L. (1993) Digital codes from hypervariable tandemly repeated DNA sequences in the Plasmodium falciparum circumsporozoite gene can genetically barcode isolates. *Molecular and Biochemical Parasitology* 61, 15–24.

Armstrong, K.F. and Ball, S.L. (2005) DNA barcodes for biosecurity: invasive species identification. *Philosophical Transactions of the Royal Society B: Biological Sciences* 360, 1813–1823.

Barbuto, M., Galimberti, A., Ferri, E., Labra, M., Malandra, R., Galli, P. and Casiraghi, M. (2010) DNA barcoding reveals fraudulent substitutions in shark seafood products: the Italian case of "palombo" (*Mustelus* spp.). *Food Research International* 43, 376–381.

Bergsten, J., Bilton, D.T., Fujisawa, T., Elliott, M., Monaghan, M.T., Balke, M, Hendrich, L., Geijer, J., Herrmann, J., Foster, G.N. *et al.* (2012) The effect of geographical scale of sampling on DNA barcoding. *Systematic Biology* 61, 851–869.

Bonants, P., Groenewald, E., Rasplus, J.Y., Maes, M., De Vos, P., Frey, J., Boonham, N., Nicolaisen, M., Bertacini, A., Robert, V. *et al.* (2010) QBOL: a new EU project focusing on DNA barcoding of quarantine organisms. *EPPO Bulletin* 40, 30–33.

Bonants, P., Edema, M. and Robert, V. (2013) Q-bank, a database with information for identification of plant quarantine plant pest and diseases. *EPPO Bulletin* 43, 211–215.

Coenye, T., Falsen, E., Vancanneyt, M., Hoste, B., Govan, J.R., Kersters, K. and Vandamme, P. (1999) Classification of *Alcaligenes faecalis*-like isolates from the environment and human clinical samples as *Ralstonia gilardii* sp. nov. *International Journal of Systematic Bacteriology* 49, 405–13.

Deng, S.J. and Hiruki, C. (1991) Genetic relatedness between two nonculturable mycoplasmalike organisms revealed by nucleic acid hybridization and polymerase chain reaction. *Phytopathology* 81, 1475–1479.

DeSalle, R., Egan, M.G. and Siddall, M. (2005) The unholy trinity: taxonomy, species delimitation and DNA barcoding. *Philosophical Transactions of the Royal Society B: Biological Sciences* 360, 1905–1916.

Edwards, U., Rogall, T., Blöcker, H., Emde, M. and Böttger, E.C. (1989) Isolation and direct complete nucleotide determination of entire genes. Characterization of a gene coding for 16S ribosomal RNA. *Nucleic Acids Research* 17, 7843–7853.

Foissac, X., Danet, J.L., Malembic-Maher, S., Salar, P., Safářová, D., Válová, P. and Navrátil, M. (2013) Tuf and secY PCR amplification and genotyping of phytoplasmas. In: Dickinson, M. and Hodgetts, J. (eds) *Phytoplasma: Methods and Protocols*. Humana Press (Springer), New York, pp. 189–204.

Francis, C.M., Borisenko, A.V., Ivanova, N.V., Eger, J.L., Lim, B.K., Guillén-Servent, A., Kruskop, S.V., Mackie, I. and Hebert, P.D.N. (2010) The role of DNA barcodes in understanding and conservation of mammal diversity in Southeast Asia. *PLoS ONE* 5(9), e12575.

Goldstein, P.Z. and DeSalle, R. (2010) Integrating DNA barcode data and taxonomic practices: determination, discovery and description. *Bioessays* 33, 135–147.

Gundersen, D.E. and Lee I.-M. (1996) Ultrasensitive detection of phytoplasmas by nested-PCR assays using two universal primer pairs. *Phytopathologia Mediterranea* 35, 144–151.

Hanner, R. and Database Working Group, Consortium for the Barcode of Life (2009) Data standards for BARCODE records in INSDC (BRIs). BARCODE Data Standards v.2.3. Consortium for the Barcode of Life, Washington, DC. Available at: http://www.barcodeoflife.org/sites/default/files/legacy/pdf/DWG_data_standards-Final.pdf (accessed 30 October 2015).

Hebert, P.D., Cywinska, A., Ball, S.L. and deWaard, J.R. (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B* 270, 313–321.

Hodgetts, J. and Dickinson, M. (2010) Phytoplasma phylogeny and detection based on genes other than 16S rRNA. In: Weintraub, P.G. and Jones, P. (eds) *Phytoplasmas: Genomes, Plant Hosts and Vectors*. CAB International, Wallingford, UK, pp. 93–113.

Hodgetts, J., Boonham, N., Mumford, R., Harrison, N. and Dickinson, M. (2008) Phytoplasma phylogenetics based on analysis of the *secA* and 23S rRNA gene sequences for improved resolution of candidate species of '*Candidatus* Phytoplasma'. *International Journal of Systematic and Evolutionary Microbiology* 58, 1826–1837.

Hollingsworth, P.M., Graham, S.W. and Little, D.P. (2011) Choosing and using a plant DNA barcode. *PLoS ONE* 6(5), e19254.

Holmes, B.H., Steinke, D. and Ward, R.D. (2009) Identification of shark and ray fins using DNA barcoding. *Fisheries Research* 92, 280–288.

Hrcek, J., Miller, S.E., Quicke, D.L.J. and Smith, M.A. (2011) Molecular detection of trophic links in a complex insect host–parasitoid food web. *Molecular Ecology Resources* 11, 786–794.

Jinbo, U., Kato, T. and Ito, M. (2011) Current progress in DNA barcoding and future implications for entomology. *Entomological Science* 14, 107–124.

Kress, W.J. and Erickson, D.L. (2008) DNA barcodes: genes, genomics, and bioinformatics. *Proceedings of the National Academy of Sciences of the United States of America* 105, 2761–2762.

Makarova, O., Contaldo, N., Paltrinieri, S., Kawube, G., Bertaccini, A. and Nicolaisen, M. (2012) DNA barcoding for identification of '*Candidatus* Phytoplasmas' using a fragment of the elongation factor Tu gene. *PLoS ONE* 7(12), e52092.

Manghisi, A., Morabito, M., Bertuccio, C., Le Gall, L., Couloux, A., Cruaud, C. and Genovese, G. (2010) Is routine DNA barcoding an efficient tool to reveal introductions of alien macroalgae? A case study of *Agardhiella subulata* (Solieriaceae, Rhodophyta) in Cape Peloro lagoon (Sicily, Italy). *Cryptogamie, Algologie*, 31, 423–433.

Puillandre, N., Bouchet, P., Boisselier-Dubayle, M.-C., Brisset, J., Buge, B., Castelin, M., Chagnoux, S., Christophe, T., Corbari, L., Lambourdière, J. *et al*. (2012) New taxonomy and old collections: integrating DNA barcoding into the collection curation process. *Molecular Ecology Resources* 12, 396–402.

Schoch, C.L., Seifert, K.A., Huhndorf, S., Robert, V., Spouge, J.L., Levesque, C.A., Chen, W. and Fungal Barcoding Consortium (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences of the United States of America* 109, 6241–6246.

Seberg, O. and Petersen, G. (2009) How many loci does it take to DNA barcode a crocus? *PLos ONE* 4(2), e4598.

Seifert, K.A. (2009) Progress towards DNA barcoding of fungi. *Molecular Ecology Resources* 9(Suppl. 1), 83–89.

Siddall, M.E., Kvist, S., Phillips, A. and Oceguera-Figuero, A. (2012) DNA barcoding of parasitic nematodes: is it kosher? *The Journal of Parasitology* 98, 692–694.

Smart, C.D., Schneider, B., Blomquist, C.L., Guerra, L.J., Harrison, N.A., Ahrens, U., Lorenz, K.-H., Seemuller, E. and Kirkpatrick, B. (1996) Phytoplasma-specific PCR primers based on sequence of the 16S-23S rRNA spacer region. *Applied and Environmental Microbiology* 62, 2988–2993.

Valentini, A., Pompanon, F. and Taberlet, P. (2009) DNA barcoding for ecologists. *Trends in Ecology and Evolution* 24, 110–117.

van de Vossenberg, B.T.L.H., Westenberg, M. and Bonants, P.J.M. (2013) DNA barcoding as an identification tool for selected EU-regulated plant pests: an international collaborative test performance study among 14 laboratories. *EPPO Bulletin* 43, 216–228.

Wandeler, P., Hoeck, P.E.A. and Keller, L.F. (2007) Back to the future: museum specimens in population genetics. *Trends in Ecology and Evolution* 22, 634–642.

Weimers, N. and Fiedler, K. (2007) Does the DNA barcoding gap exist? – a case study in blue butterflies (Lepidoptera: Lycaenidae). *Frontiers in Zoology* 4, 8.

White, T.J., Bruns, T., Lee, S. and Taylor, J. (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis, M.A., Gelfand, D.H., Sninsky, J.J. and White, T.J. (eds) *PCR Protocols: A Guide to Methods and Applications*. Academic Press, San Diego, California, pp. 315–322.

Will, K.W., Mishler, B.D. and Wheeler, Q.D. (2005) The perils of DNA barcoding and the need for integrative taxonomy. *Systematic Biology* 54, 844–581.

Yancy, H.F., Zemlak, T.S., Mason, J.A., Washington, J.D., Tenge, B.J., Nguyen, N-L.T., Barnett, J.D., Savary, W.E., Hill, W.E., Moore, M.M. *et al*. (2008) Potential use of DNA barcodes in regulatory science: applications of the *Regulatory Fish Encyclopedia*. *Journal of Food Protection* 71, 210–217.

6 DNA barcoding of invertebrate plant pests

Sebastian Kiewnick,* Andreas Bühlmann and Jürg E. Frey

Agroscope, Wädenswil, Switzerland

6.1 Introduction

The term barcoding in the context of species identification was first used in 1993 when Arnot and co-workers presented a study on different *Plasmodium falciparum* isolates (Arnot *et al.*, 1993). In the following years, much effort went into exploring suitable genetic markers for species allocation, and soon the mitochondrial cytochrome oxidase I gene was among the preferred genetic markers (e.g. Brunner *et al.*, 2002). Almost a decade later, DNA barcoding emerged as a new discipline in biological research (Hebert *et al.*, 2003b). Several consortia and databases where established by international efforts, with the Consortium for the Barcode of Life (CBOL; http://www.barcodeoflife.org) and the Barcode of Life Data Systems (BOLD; http://www.boldsystems.org) being the most well known. While the debate about the feasibility of barcoding in describing new species is still ongoing (Rubinoff *et al.*, 2006; Taylor and Harris, 2012), its main use – in species identification and recognition – is not subject to debate. In particular, for invertebrates, for which morphological identification can be challenging owing to their morphologically different live stages (Ahrens *et al.*, 2007), rapid adoption of the technology and a great number of publications and reference sequences in the databases (Dupuis *et al.*, 2012; Taylor and Harris, 2012) could be observed. In plant pest diagnostics, arthropods (Brunner *et al.*, 2002; Frey *et al.*, 2004, 2013; Frey and Pfunder, 2006), mites (Ros and Breeuwer, 2007) and nematodes (Floyd *et al.*, 2005) have seen great advances in DNA barcoding over the past decade.

*Corresponding author: sebastian.kiewnick@agroscope.admin.ch

Although the barcoding gene of choice, that for mitochondrial cytochrome c oxidase subunit I (*COI*) (Hebert *et al.*, 2003b) initially raised concerns because of its mitochondrial origin (Rubinoff and Holland, 2005), and has shown to be troublesome in some barcoding initiatives such as that of benthic cnidarians (Hebert *et al.*, 2003a), amphibians (Vences *et al.*, 2005) and gastropods (Remigio and Hebert, 2003), it has been valuable for the identification of most arthropod plant pests (Brunner *et al.*, 2002; Frey *et al.*, 2013). The mitochondrial origin renders the gene free from introns and recombination events, thereby simplifying analysis (Waugh, 2007). The sequencing of *COI*, especially the 648 bp long 5′ region termed the barcoding fragment (Hebert *et al.*, 2003b), has been shown to be a valuable approach for the identification of lepidopteran (Hajibabaei *et al.*, 2006) and dipteran pests (Meier *et al.*, 2006), whereas nematodes and fungi need additional or other sequences such as those of 18S rRNA or the intergenic transcribed spacer (ITS) (Floyd *et al.*, 2005; Seifert, 2009). Due to its relatively low costs and objective nature, barcoding adds value to the identifications of classical taxonomists or relieves taxonomists from routine work (Valentini *et al.*, 2009). This is especially true when gathering morphological information is time-consuming, difficult or even impossible as the organisms are damaged, there are different morphological life stages (Ahrens *et al.*, 2007), or species cannot be discriminated by using morphological or morphometric features (Saunders, 2005). Furthermore, specimens with DNA barcodes, vouchered in reliable databases, may be reanalysed in the future, e.g. for population genetics studies or the detection of invasive species (Valentini *et al.*, 2009). Beyond serving as a heuristic in identifying specimens of described species, barcoding may, though this is much debated, assist in generating hypotheses of new and/or cryptic species (Hajibabaei *et al.*, 2007; Goldstein and DeSalle, 2011), thus serving an integrated taxonomy approach (Will *et al.*, 2005).

The analysis of DNA barcode sequences has been done by simple BLAST (the Basic Local Alignment Search Tool from GenBank, the US National Institutes of Health genetic sequence database available from NCBI, the US National Center for Biotechnology Information) analysis (Altschul *et al.*, 1990), using distance-based phylogenetic methods such as neighbour-joining trees, a method that has its drawbacks (Lipscomb *et al.*, 2003; DeSalle *et al.*, 2005), but is computationally simple and fast (Saitou and Nei, 1987). BOLD uses a neighbour-joining approach (Ratnasingham and Hebert, 2007) that relies on the so-called barcode gap (the gap between intraspecific variation and interspecific divergence in the selected marker), which can be missing in some species (Meyer and Paulay, 2005) and can therefore cause problems in identification within some taxa. Other approaches are based on likelihood ratio tests (Matz and Nielsen, 2005) or a Bayesian analysis (Nielsen and Matz, 2006). Recently, character-based methods have been refined to circumvent the barcode gap problem (Sarkar *et al.*, 2008).

In this chapter, we will summarize the current state-of-the-art in barcoding for invertebrate plant pests, focusing on insects, mites and nematodes, and also provide a basic protocol for DNA barcode sequencing and data analysis.

6.2 Barcoding nematodes

The Nematoda constitute a taxon for which the majority of the diversity has not yet been described, so there is a significant identification gap (Powers *et al.*, 2011). This lack of described species has mainly been attributed to several facts: nematodes are

small in size; they have high intraspecific variation combined with a simple morphology; and the number of expert nematode taxonomists is declining. The circumstance that often up to 90% of the specimens in a given sample cannot be assigned to a known species has spurred research into the use of molecular approaches to improve the discovery of species in biodiversity studies.

The early beginnings of the development of molecular fingerprints for nematode species identification utilized a number of methods, such as restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD) analysis and amplified fragment length polymorphism (AFLP) (Floyd *et al.*, 2002; Blaxter, 2004; Powers, 2004). However, although yielding barcode-like data, these methods were not optimal for molecular nematode taxonomy because the within-taxon variability was too high (Blaxter, 2004).

The use of DNA barcodes offers a unique approach in sequence-based molecular taxonomy. It allows the utilization of universal protocols for DNA extraction, PCR amplification and sequencing. Moreover, high-throughput sequencing can be used to enable even non-experts in nematode taxonomy to identity a vast number of species correctly. The universal protocol can be used with a single specimen, independent of its life stage, whereas the different life stages (juvenile, female, male) are required for morphological or biochemical identification.

6.2.1 Genes and gene regions for molecular barcoding studies (biodiversity)

Initially, the DNA barcoding of nematodes was used in studies on taxonomy and biodiversity. Research focused on areas such as marine nematodes, with an expected one million species, of which only a few thousand had been described (Bhadury *et al.*, 2006), or on soil nematode communities with several hundred described species obtained from a single sampling site (Floyd *et al.*, 2002). Candidate DNA barcode sequence targets have to fulfil certain requirements, such as being orthologous and showing sufficient variability; this will allow the required discrimination between taxa (Blaxter, 2004). The overall goal is to achieve a separation into molecular operational taxonomic units (MOTUs) (Floyd *et al.* 2002; Blaxter, 2004). MOTUs have previously been named phylotypes or genospecies (Blaxter, 2004), and are defined by sequence identity within a predefined cut-off value. It should be noted that assigning a specimen to a certain MOTU does not correspond to placing it into a taxonomic unit defined by morphological or biological features (Blaxter, 2004); the taxonomic significance of a sequence depends on several factors, such as the chosen gene or gene region, the evolution of this particular region, potential experimental errors and whether the results correspond to phylogenetic trees (Powers *et al.*, 2011).

6.2.2 Choice of barcode regions

Initially, barcode studies on biodiversity in nematology focused on the nuclear small-subunit RNA (SSU) and the large-subunit ribosomal RNA gene (*LSU*), as well as the highly variable internal transcribed spacer region of the ribosomal RNA cistron (ITS1

and ITS2). Most studies made use of the SSU and LSU regions, as a significant number of sequences had been collected for these and were available for extensive BLAST match searches (De Ley *et al.*, 2005; Holterman *et al.*, 2006; Porazinska *et al.*, 2009; Van Megen *et al*, 2009). Furthermore, primer sites for these target regions are relatively conserved across nematode clades and sequence data allowed sufficient resolution to identify specimens to species level (Porazinska *et al.*, 2009).

Researchers have also explored the *COI* gene for the DNA barcoding of nematodes in biodiversity studies, with promising results (Hu *et al.*, 2002; Powers, 2004; Derycke *et al.*, 2005, 2006, 2007 and 2010 (and references therein); Ye *et al.*, 2007). These studies indicated that this gene might not be as conserved in nematodes and could, therefore, allow good species discrimination.

The use of barcoding for the identification of quarantine plant parasitic nematodes regulated by the European Union (EU) Council Directive 2000/29/EC and their closely related species has been approached only recently (Bonants *et al.*, 2010). The molecular diagnostics of quarantine nematodes has mostly focused on the nuclear ribosomal DNA (rDNA) cistron, which consists of coding and non-coding regions. The relatively short non-coding regions have allowed discrimination between related species of plant parasitic nematodes (Zijlstra *et al.*, 1995; Blok *et al.*, 1997). Some diagnostic protocols have exploited the unique characteristics of the mitochondrial genome, but its use as a barcoding region on a larger scale was not envisaged at that time (Powers, 2004). With the development of high-throughput DNA sequencing, larger and moderately variable coding regions, such as the SSU ribosomal RNA gene (Blaxter *et al.*, 1998; Tandingan De Ley *et al.*, 2002; Holterman *et al.*, 2006; Van Megen *et al.*, 2009) were chosen for sequence-based diagnostics. The *SSU* gene, although generally considered to be highly conserved, has been shown to be more variable in nematodes, in many cases allowing discrimination down to species level (Holterman *et al.*, 2009). Other studies focused on the D1–D2 and D2–D3 regions of the *LSU* gene (Subbotin *et al.*, 2005; Holterman *et al.*, 2008; Landa *et al.*, 2008) and demonstrated a higher variability of this gene when it was compared with the *SSU* gene. The ITS regions of the ribosomal RNA cassette have also been used to discriminate between species, but this non-coding region contains more within-species variation than the *SSU* and *LSU* genes, which makes it unsuitable for barcoding (Subbotin *et al.*, 2000; Espineira *et al.*, 2010; Madani *et al.*, 2010; Marek *et al.*, 2010; Maneesakorn *et al.*, 2011).

Within the EU 7th Framework project QBOL (Development of a New Diagnostic Tool Using DNA Barcoding to Identify Quarantine Organisms in Support of Plant Health; Project No. KBBE-2008-1-4-01), the overall goal was to develop barcoding as a tool for identifying regulated quarantine organisms of importance for Europe (http://www.qbol.org; Bonants *et al.*, 2010). This included establishing DNA barcodes for all regulated quarantine organisms and their close relatives, followed by the development of the curated, freely accessible DNA barcode database, Q-bank, a European initiative which is a collection of comprehensive databases on quarantine plant pests and diseases (www.Q-bank.eu) containing the reference DNA/RNA sequences of Q-organisms linked to digital vouchers and taxonomic metadata. In particular, only correctly identified

material from reference collections was used to generate high-quality sequences data deposited in the database, so that the barcode sequences could be used as an identification tool for national plant protection organizations.

To achieve these goals for nematodes, candidate barcoding regions were selected and evaluated in the initial proof-of-concept phase (Holterman *et al.*, 2011). In a first step, the most important quarantine species and closely related species were placed in priority group 1 (Table 6.1). Group 1 was used for the proof-of-concept; the candidate genes/gene with good barcode regions were considered, along with the protocols developed for DNA

Table 6.1. List of quarantine nematode species and closely related species used for a tiered approach in developing barcoding as an identification tool.

Group 1	Group 2	Group 3	
Bursaphelenchus chengi	*Aphelenchoides besseyi*[a]	*Anguina tritici*	*Longidorus macrosoma*
Bursaphelenchus doui	*Ditylenchus destructor*	*Aphelenchoides bicaudatus*	*Meloidogyne exigua*
Bursaphelenchus mucronatus	*Ditylenchus dipsaci*[a]	*Aphelenchoides ritzemabosi*	*Paratrichodorus anemones*
Bursaphelenchus thailandae	*Globodera pallida*[a]	*Aphelenchoides saprophilus*	*Paratrichodorus porosus*
Bursaphelenchus xylophilus[a]	*Globodera rostochiensis*[a]	*Bursaphelenchus conicaudatus*	*Paratrichodorus pachydermus*
Meloidogyne arenaria	*Heterodera glycines*[a]	*Bursaphelenchus fraudulentus*	*Pratylenchus crenatus*
Meloidogyne chitwoodi[a]	*Hirschmanniella* spp.[a]	*Ditylenchus adasi*	*Pratylenchus mediterraneus*
Meloidogyne enterolobii[a]	*Longidorus diadecturus*[a,b]	*Globodera achilleae*	*Pratylenchus loosi*
Meloidogyne fallax[a]	*Meloidogyne ardenensis*	*Globodera tabacum*	*Pratylenchus neglectus*
Meloidogyne hapla – race A	*Meloidogyne maritima*	*Heterodera avenae*	*Pratylenchus penetrans*
Meloidogyne incognita	*Nacobbus aberrans*[a]	*Heterodera filipjevi*	*Pratylenchus thornei*
Meloidogyne javanica	*Radopholus similis*[a] (attacking citrus; formerly *R. citrophilus)*	*Heterodera latipons*	*Pratylenchus vulnus*
Meloidogyne minor	*Radopholus similis* (not attacking citrus)	*Heterodera schachtii*	*Radopholus arabocoffeae*
Meloidogyne naasi	*Xiphinema bricolense*[a]	*Helicotylenchus dihystera*	*Radopholus duriophilus*

Continued

Table 6.1. Continued.

Group 1	Group 2	Group 3	
Xiphinema americanum sensu lato[a]	*Xiphinema californicum*[a]	*Helicotylenchus pseudorobustus*	*Subanguina radicicola*
Xiphinema diversicaudatum		*Helicotylenchus varicaudatus*	*Trichodorus cedarus*
Xiphinema rivesi[a]		*Helicotylenchus vulgaris*	*Trichodorus cylindricus*
		Hirschmanniella gracilis	*Trichodorus primitivus*
		Laimaphelenchus penardi	*Xiphinema incognitum*
		Longidorus arthensis	*Xiphinema index*
		Longidorus dunensis	*Xiphinema simile*

[a]Regulated species (for details, see www.eppo.org).
[b]Only one record for Canada, no material available.

extraction, and these were then tested with the species listed in group 2. For the species listed in priority group 3, additional species reference material was obtained (Table 6.2), and sequences were generated based on the outcome of the proof-of-concept phase, following a generic barcoding protocol.

For the proof-of-concept, the *SSU* rRNA gene, a 1000 bp region of the *LSU* rRNA gene, the *COI* and *COII* (cytochrome c oxidase subunit II) genes and a 1000 bp fragment of the ribosomal polymerase 2 gene were chosen. Primers were taken either from the literature or new primers were designed to meet the purpose. For some species in priority group 1, the intergenic spacer region 2 (IGS2) was tested for barcode capabilities, but initial results showed that within-species variation was too high (Kiewnick *et al.*, 2011).

6.2.3 Barcoding as an identification tool

Intensive testing of DNA barcoding regions revealed that, as expected, universal primers cannot be used for all candidate genes. For example, the standard *COI* primers (Derycke *et al.*, 2010) failed to produce amplicons with *Ditylenchus dipsaci* but worked well with all other species tested (Tables 6.1 and 6.2). Newly designed *COII* primers worked with a wide range of nematode species, including the genera *Meloidogyne*, *Globodera*, *Heterodera*, *Helicotylenchus*, *Rotylenchus*, *Radopholus*, *Hirschmanniella*, *Nothotylenchus*, *Bursaphelenchus* and *Aphelenchoides*, but failed to produce amplicons with *Pratylenchus*, *Ditylenchus*, *Anguina*, *Ecphyadophora*, *Hemicriconemoides*, *Hemicycliophora*, *Macrotrophurus*, *Nagelus*, *Neodolichorhynchus* and *Xiphinema* species.

Table 6.2. List of additional nematode species used to validate generic barcoding protocols and gene regions.

Species		
Aphelenchoides blastophthorus	*Helicotylenchus lobus*	*Nagelus obscurus*
Aphelenchoides cf. bicaudatus	*Helicotylenchus microcephalus*	*Neodolichorhynchus lamellipherus*
Aphelenchoides cf. breviutoralis	*Helicotylenchus minzi*	*Neodolichorhynchus microphasmis*
Aphelenchoides composticola	*Helicotylenchus phalerus*	*Nothotylenchus acris*
Aphelenchoides fragariae	*Helicotylenchus pseudodigonicus*	*Ogma menzeli*
Aphelenchoides subtenuis	*Hemicriconemoides pseudobrachyuris*	*Pratylenchus pratensis*
Bitylenchus dubius	*Hemicycliophora thienemanni*	*Pratylenchus scribneri*
Bursaphelenchus eremus	*Longidorus attenuatus*	*Pseudhalenchus minutus*
Bursaphelenchus paraluxuriosae	*Macrotrophurus arbusticola*	*Rotylenchulus borealis*
Diphtherophora communis	*Meloidogyne artiellia*	*Rotylenchus fallorobustus*
Ditylenchus angustus	*Meloidogyne duytsi*	*Rotylenchus goodeyi*
Ditylenchus feropoliter	*Meloidogyne ethiopica*	*Rotylenchus reniformis*
Ecphyadophora tenuissima	*Meloidogyne graminicola*	*Rotylenchus robustus*
Halenchus fucicola	*Meloidogyne hispanica*	*Rotylenchus uniformis*
Helicotylenchus canadensis	*Meloidogyne ichinohei*	*Sauertylenchus maximus*
Helicotylenchus crenacauda	*Meloidogyne kralli*	*Trophurus imperialis*
Helicotylenchus digitiformis	*Meloidogyne mali*	*Tylenchorhynchus teeni*
Helicotylenchus digonicus	*Meloidogyne maritima*	*Tylolaimophorus typicus*
Helicotylenchus egyptiensis	*Meloidogyne thailandica*	*Xiphinema chambersi*
Helicotylenchus exallus	*Meloidogyne ulmi*	
Helicotylenchus indicus	*Mesocriconema xenoplax*	

With the development of barcoding for nematode identification, it became clear that a proper molecular tool for identifying nematode species should consider more than one gene or gene region (Blaxter, 2004; Derycke *et al.*, 2010). As the key deliverable of the QBOL project, more than 1600 sequences, representing 120 nematode species, were deposited in the Q-bank database. The unique feature of this database is that next to the use of single locus BLAST searches, multilocus searches can be conducted. Using this approach, 120 species for which sequences were generated can be identified using two sequences obtained following a generic barcoding protocol. For a robust and reliable identification of quarantine nematode species and their close relatives, at least two unrelated barcoding genes, such as *COI* and *SSU* rDNA, should be combined within the DNA barcoding system (see Fig. 6.1).

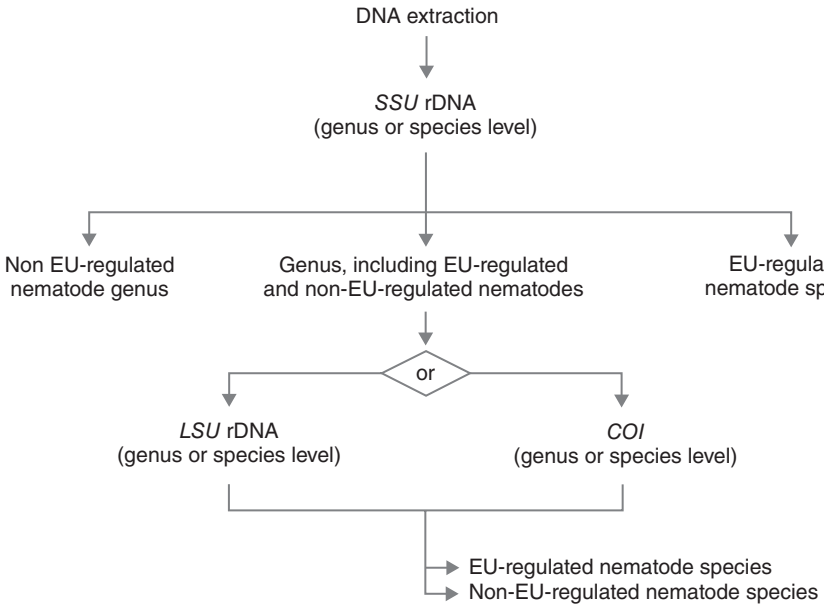


Fig. 6.1. Molecular decision scheme for the identification of quarantine nematode and closely related species by barcoding. (From www.Q-bank.eu/Nematodes.) Key: COI, mitochondrial cytochrome c oxidase subunit I; LSU, large subunit rRNA; SSU, small subunit rRNA.

6.2.4 Outlook for barcoding and identification

With the establishment of databases in which thousands of sequences have been deposited, mostly combined with metadata such as digital and DNA vouchers, the backbone for the identification of widespread known nematode species, and also for that of new and emerging species, has been built. This backbone of barcode sequences has increasingly been used for nematode identification and the description of new species (Gutiérrez-Gutiérrez *et al.*, 2012; de Luca *et al.*, 2013; Vovlas *et al.*, 2013). In addition, these barcode sequences have served as a source for the development of new, highly specific and sensitive assays for the identification of plant parasitic nematodes, including quarantine species, within complex DNA backgrounds (Rybarczyk-Mydłowska *et al.*; 2012; Kiewnick *et al.*, 2013; Toumi *et al.*, 2013).

6.3 Barcoding arthropods

With an estimated 5–10 million described species, arthropods account for well over 80% of all known living animal species (Odegaard, 2000). It is, therefore, no surprise that considerable effort has been put into developing DNA barcoding for this important animal group. As one indicator of the recent advances in DNA barcoding, searches of the PubMed database for the terms 'barcoding' and/or 'species identification' give a list over 1000 citations for arthropods alone. Barcoding has been used with great success in studies on biodiversity estimation (e.g. Janzen and Hallwachs, 2011; Ji *et al.*, 2013), community ecology (e.g. Tanzler *et al.*, 2012), the discovery of cryptic species

(e.g. Hebert *et al.*, 2004), host–parasite biology (e.g. Brunner *et al.*, 2004; Smith *et al.*, 2009) and species distribution and phenology (e.g. Brunner and Frey, 2010; deWaard *et al.*, 2010). Recently, overviews have been published on the (general) use of barcoding for arthropods (e.g. Hajibabaei *et al.*, 2005; Evans and Paulay, 2012; Wilson, 2012) and in studies on specific pest groups, e.g. tephritid fruit flies (Frey *et al.*, 2013), storage pest psocids (Yang *et al.*, 2013), bugs (Park *et al.*, 2011a; Rebijith *et al.*, 2012), aphids (Lee *et al.*, 2011; Chen *et al.*, 2012) and scales (Park *et al.*, 2011b). There have also been studies published on the use of barcoding for small groups or individual species of important and often invasive pests (e.g. deWaard *et al.*, 2010; Nagoshi *et al.*, 2011; Sutou *et al.*, 2011; Khamis *et al.*, 2012). Furthermore, information and detailed protocols on barcoding have been made available on various Internet resources, such as Cold Spring Harbor Laboratory's Barcoding 101 (at http://www.dnabarcoding101.org/), the Barcoding of Life Database resources (BOLD, at http://www.boldsystems.org/) and the Canadian Centre for Barcoding site (CCDB, at http://www.ccdb.ca/).

The majority of barcoding efforts for invertebrates used the first *c.*700 bp of the mitochondrial *COI* gene, which now represents the standard barcoding marker (Hebert *et al.*, 2003b). Particularly for the DNA barcoding of arthropods, this fragment has proven to be an excellent choice, as there is in general a significantly higher (on average 10×) differentiation among species than within species. In a few cases in which the level of differentiation of this standard barcode did not allow for discrimination of all morphological species in a system (e.g. Frey *et al.*, 2013), parts of other genes were successfully used either as alternative markers or in combination with *COI*. Suitable combinations were with other mitochondrial genes, such as *COII* and the cytochrome b, 16S rRNA, NADH dehydrogenase 1 genes, as well as the multicopy nuclear ribosomal genes (e.g. the ITS or the 28S portion of the ribosomal rRNA genes), or the single-copy nuclear gene elongation factor 1α (e.g. Monaghan *et al.*, 2005; Lefebure *et al.*, 2006; Elias *et al.*, 2007; Foley *et al.*, 2007; Rach *et al.*, 2008; Acs *et al.*, 2010; Damm *et al.*, 2010; Kim *et al.*, 2010; Leo *et al.*, 2010; Thormann *et al.*, 2011). Other single-copy nuclear genes such as arginine kinase (Hawlitschek *et al.*, 2011), lysozyme (Leo *et al.*, 2010), or acetylcholine esterase 2 (Hemmerter *et al.*, 2009) have been used occasionally. One notable exception from the regular use of the *COI* barcoding region for the reliable identification of a pest species is the whitefly, *Bemisia tabaci* (Hemiptera, Aleyrodidae), in which, due to the limited discrimination power of the barcode fragment, the use was proposed of a region more to the 3′-end of the *COI* gene (Shatters *et al.*, 2009).

A prerequisite for the DNA barcoding of a diverse animal group such as arthropod pests is successful PCR amplification. With the high-quality chemistry available today, the main challenge lies in the sequence of the primers. Details of the primers needed for PCR amplification and sequencing of the genes mentioned above are given in the papers cited and in compilations for mitochondrial genes (Folmer *et al.*, 1994; Simon *et al.*, 1994, 2006), or for nuclear primers in Hymenoptera in Hartig *et al.* (2012). The BOLD database also provides primer sequences. In addition, Park *et al.* (2010) developed primers that should amplify and enable sequencing of the relevant *COI* gene regions to design *de novo* barcoding primers for all arthropods.

One challenge in the barcoding of arthropods is obtaining a high-quality DNA sequence of the targeted barcode fragment. Although primers for this purpose have been optimized, arthropods present a large and genetically highly diverse group of animals, which means that the requirements for correct species identification can be conflicting. Today, one important application for DNA barcoding is to identify the arthropods intercepted at ports of entry, such as airports. Usually, there is a suspicion about the genus or family to which a specimen may belong, but this still leaves many possible species that have to be differentiated. Therefore, the first requirement is that the primers used are able to amplify the barcoding fragment of all species of a genus or even a family. This can be achieved by using degenerate primers and/or by compensating for primer mismatch with decreasing annealing temperatures. However, this approach is in direct conflict with the second requirement, which advocates use of the highest possible stringency of the system to produce clean amplicons but, at the same time, minimize the risk of accidental co-amplification of alternative genetic targets. Such alternative target sequences are known to occur in most arthropod species and include mainly nuclear introgressed mitochondrial pseudogenes, the so-called NUMTs (Zhang and Hewitt, 1996; Moulton *et al.*, 2010). Additionally, the presence of mixed haplotypes caused by heteroplasmy may hinder successful identification (Boyce *et al.*, 1989; Frey and Frey, 2004; Magnacca and Brown, 2010). These factors can interfere with diagnosis, as they can lead to co-amplification of false templates, in particular when frequent primer mismatches are involved in barcoding highly diverse arthropod pests.

Sanger sequencing (chain termination dideoxy sequencing) is currently the standard method; here, the resulting DNA sequence represents the consensus sequence of all amplicons produced during PCR. If co-amplification of NUMTs and/or heteroplasmy occurs in a specimen, the resulting consensus sequence will not be affected unless the fraction of the second most common target in the amplicon mixture is well above 30% (Frey and Frey, 2004). Levels of 30% or less will appear on the sequence chromatogram as noise in the form of minor secondary peaks under the correct ones. The distribution of both phenomena is very stochastic, even among related species (Song *et al.*, 2008). In heteroplasmy, the frequency of alternative targets is typically below the threshold that would confound the consensus sequence (Frey and Frey, 2004). Dealing with NUMTs can be more challenging (Song *et al.*, 2008) because the longer the time since nuclear introgression, the more likely the target will have accumulated mutations, including insertions and/or deletions leading to stop codons. These may easily be identified and hence do not present problems, whereas more recently introgressed sequences may not contain clear signatures, leading to misinterpretation, and thereby hindering the barcoding process. However, most nuclear introgressed mitochondrial fragments are less than approximately 350 bp in length and so too short to be misidentified as a correct template (Pamilo *et al.*, 2007). Furthermore, although multiple copies of the same NUMT can be present, their number rarely reaches the larger number of mitochondrial genes that are generally encountered in single cells (Frey and Frey, 2004). Unless the primers have a significantly better

match with the NUMT template than with the targeted mitochondrial template, only a very limited amplification of full-length false-positive templates can occur (Frey and Frey, 2004).

Taking into account the above-stated considerations, and with a focus on the simplicity and robustness of methods and processes, a *COI*-based DNA barcoding strategy (see Fig. 6.2) was developed in our laboratory that enabled the identification of *c.*150 arthropods of ten different families to the species level (Table 6.3). The same barcoding strategy was used in the EU project QBOL and delivered an additional 18 species of the taxonomically difficult family Tephritidae (Diptera), which contains many important quarantine species (Frey *et al.*, 2013). The consequence of such a broad taxonomic coverage is that the genetic variation among taxa at least partly affects primer sites as well. Therefore, protocols have to be developed that on one hand enable PCR amplification with primers that have mismatches, but on the other hand are, in general, stringent enough to inhibit amplification of possible alternative target sequences from other genes. Under these conditions, the most important aspects relevant for successful PCR are: (i) the DNA extraction method; (ii) the specific *Taq* polymerase (from *Thermus aquaticus*) used; and (iii) the PCR thermocycling programme. These factors are strongly interdependent and need to be fine-tuned to establish a robust amplification system. This system only requires a simple and low-cost DNA extraction, one single PCR protocol and a limited set of just six primer pairs, and successfully amplify more than 90% of these taxa.

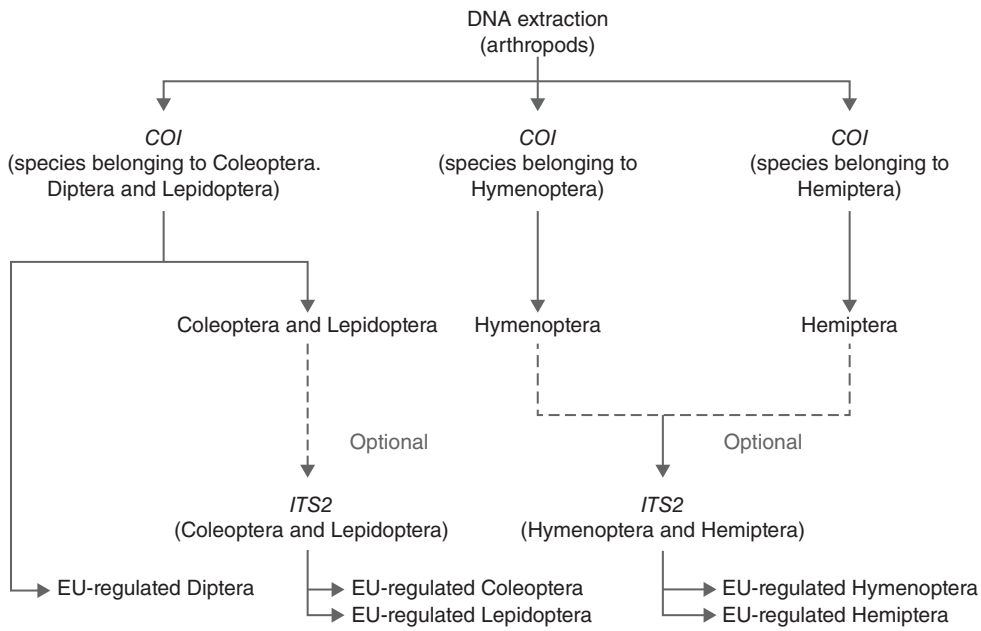


Fig. 6.2. Molecular diagnostic decision scheme for identification of quarantine arthropod species and closely related species. (From www.Q-bank.eu/Insects.) Key: COI, mitochondrial cytochrome c oxidase subunit I; ITS, intergenic spacer region 2.

Table 6.3. List of diagnosed arthropod species. For primer sets, see Table 6.4.

Order	Family	No. of genera	No. of species	PCR programme[a]	Primer set
Acarina	Laelapidae	1	2	1	1 + 2, 3 + 4
	Phytoseiidae	3	5	1	1 + 2
	Tetranychidae	1	1	1	3 + 4
Blattodea	Blattidae	1	1	1	1 + 2
Coleoptera	Coccinellidae	5	5	1	1 + 2, 3 + 4, 5 + 6
	Curculionidae	3	4	1	1 + 2, 3 + 4, 3 + 9
	Tenebrionidae	1	1	1	1 + 2
Diptera	Agromyzidae	2	4	1	1 + 2, 3 + 9
	Anthomyiidae	1	3	1	3 + 4
	Cecidomyiidae	2	4	1	3 + 4
	Drosophilidae	2	3	1	16 + 17
	Muscidae	2	2	1	16 + 21, 1 + 2, 16 + 17, 3 + 9
	Phoridae	1	1	1	16 + 21, 1 + 2
	Tephritidae	7	36	1	17 + 18, 18 + 9, 3 + 4
Hemiptera, Aphidoidea	Aphididae	7	12	1	1 + 2, 5 + 8, 3 + 4
Hemiptera, Auchenorrhyncha	Cercopidae	1	1	1, 2	5 + 8
	Delphacidae	1	1	1	5 + 8, 3 + 4
Hemiptera, Heteroptera	Anthocoridae	2	4	1	1 + 2, 3 + 4, 5 + 8
	Miridae	2	2	1	1 + 2
	Pentatomidae	1	1	1	1 + 2
Hemiptera, Homoptera	Diaspididae	3	5	1	5 + 6, 7 + 6
Hemiptera, Sternorrhyncha	Aleyrodidae	2	2	1	19 + 20, 19 + 21
	Psyllidae	2	2	1	1 + 2, 3 + 4, 5 + 8
Hymenoptera	Aphelinidae	4	5	1	1 + 2, 3 + 4
	Braconidae	3	4	1	1 + 2
	Encyrtidae	3	3	1	1 + 2
	Eulophidae	1	1	1	1 + 2
	Trichogrammatidae	1	2	1	1 + 2, 3 + 4

Continued

Table 6.3. Continued.

Order	Family	No. of genera	No. of species	PCR programme[a]	Primer set
Lepidoptera	Crambidae	2	2	1	1 + 2
	Noctuidae	3	4	1	1 + 2, 3 + 4
	Plutellidae	1	1	1	1 + 2
	Pyralidae	1	1		3 + 4
	Tortricidae	5	6	1	1 + 2
Orthoptera	Acrididae	1	1	1	3 + 4
Thysanoptera	Thripidae	12	18	1, 2	3 + 4, 10 + 11, 12 + 13, 14 + 15

[a]Programme 1, Initial denaturation: 15 min at 95°C, 40–45 cycles with 40 s denaturation at 95°C, 15 s annealing at 45°C, temperature ramping to 60°C at 15°C per min, 5 s hold at 60°C, 2 min extension at 72°C; 15 min final extension at 72°C. Programme 2, 15 min at 95°C, 45 cycles with 45 s denaturation at 95°C, 45 s annealing at 50°C, 60–90 s extension at 72°C; 7 min final extension at 72°C.

Table 6.4. Primers used for arthropod barcoding (as listed in Table 6.3).

Primer no.	Name	Sequence[a]	Source
1	LCO1490	GGTCAACAAATCATAAAGATATTGG	Folmer *et al.*,1994
2	HCO2198	TAAACTTCAGGGTGACCAAAAAATCA	
3	Ron (C1J1751)	GGATCACCTGATATAGCATTCCC	Simon *et al.*, 1994
4	Nancy (C1N2191)	CCCGGTAAAATTAAAATATAAACTTC	
5	C1J1718 (mtd6)	GGAGGATTTGGAAATTGATTAGTTCC	
6	Nancy-Scale	GGTAAAATTAAAATATAWACTTCWGG	
7	C1N1560reverse	TGAGCCGGAATAGTAGGAACA	Modified from Simon *et al.*,1994
8	C1N2329 (mtd11)	ACTGTAAATATATGATGAGCTCA	Simon *et al.*, 1994
9	C1N2353	GCTCGTGTATCAACGTCTATWCC	Simon *et al.*, 2006
10	mtd-7.2F	ATTAGGAGCHCCHGAYATAGCATT	Brunner *et al.*, 2002
11	mtd-9.2R	CAGGCAAGATTAAAATATAAACTTCTG	
12	Thys_COI_degF1	GAHGGDGCGGAACDGGDTGAA	Timm *et al.*, 2008
13	Thys_COI_degR1	GGRTCHCCWCCTCCTCYHGGRTCAA	
14	Thys_CO1_IntDegF1	GGDATHTCHTCWATYYTAGGDGC	
15	Thys_COI_IntDegR1	GCHCCTARRATWGADGADAYHCC	
16	TY-J-1460mod	TACARTCTATYGCCTAAACTTCAGC	Frey *et al.*, 2013
17	TL2-N3014mod	CATTGCACTAWTCTGCCATATTAG	

Continued

Table 6.4. Continued.

Primer no.	Name	Sequence[a]	Source
18	C1-J-1751	GGATCACCTGATATAGCATTCCC	Simon *et al.*, 1994
19	C1J2195	TTGATTTTTTGGTCATCCAGAAGT	
20	2115-R	CCAGGAACAGAATCAACCTTTACC	Shatters *et al.*, 2009
21	TL2N3014	TCCAATGCACTAATCTGCCATATTA	Simon *et al.*, 1994

[a]A, adenine; C, cytosine; G, guanine; T, thymine.

6.4 Conclusions and outlook

One of the main goals of the recent studies on the evaluation of the DNA barcoding of arthropods and nematodes was to provide National Plant Protection Organizations (NPPOs) with a new tool for the identification of non-indigenous or quarantine species encountered. Floyd *et al.* (2010) states that DNA barcoding fulfils the requirements to be included in the diagnostic protocols under the International Standards for Phytosanitary Measures (ISPM) No. 27, Diagnostic Protocols for Regulated Pests. Barcoding protocols are flexible, sensitive and specific, allow for reproducibility and are subject to review and amendments.

Importantly, the success of DNA barcoding depends on strict quality control and a high level of maintenance of the existing databases to ensure that sequences from all relevant (regulated and close relative) species are included that have the required quality and are accompanied by digital and DNA vouchers from acknowledged phytosanitary collections (Bonants *et al.*, 2010).

Protocols

Contents

Protocol 6.1

Generic DNA barcoding protocol for arthropods and nematodes

Method

1. Sampling and storage

Arthropods
Sampling for arthropods is mostly done at airports where commodities are selected randomly from shipments and checked by the appropriate methods for the presence of live arthropods, with particular reference to putative pests of quarantine origin. If arthropods are found, pictures are taken from a subset of the samples, and specimens are then transferred into 70% ethanol (EtOH) and shipped via express mail to the diagnostic laboratory.

Nematodes
Sampling is done according to national regulations from consignments with rooted plants and plants for planting. Soil and/or root samples are taken and sent to a diagnostic laboratory equipped for soil nematode extraction. Nematode suspensions are then checked for the presence of suspicious species and specimens are collected for DNA extraction.

2. DNA extraction

Arthropods
Specimens from the EtOH are transferred to a deep well tube containing a proteinase K buffer and a sterile stainless steel ball bearing. The entire sample is to be used when the body size is 3 mm or less (e.g. thrips, leaf miners, whiteflies, spider mites). For larger specimens, only parts of the bodies are taken (e.g. one leg or one antenna from Lepidoptera, Diptera and others, or a small part of the body of caterpillars or maggots). The buffer volume is adjusted to the sample size; with very small samples (thrips, aphid eggs, etc.), only 50 µl proteinase K buffer is used, mid-sized samples are homogenized with 100 µl buffer and larger samples with 200 µl buffer. The samples are briefly heated and then homogenized in a tissue homogenizer. After brief centrifugation, 1–3 µl of the supernatant is directly used as the PCR template. This rapid DNA preparation method works well with most arthropods, with the notable exception of aphids,

in which the crude DNA template inhibits PCR, probably as a result of high sugar contents. Therefore, the crude DNA template obtained from juvenile and adult stages of aphids must be further purified using a commercial silica column-based system.

Nematodes

DNA extraction is a critical step in the DNA barcoding process of nematodes. Several protocols have been developed, mostly based on a lysis buffer to extract the minute amounts of DNA from specimens. These protocols are described in detail in Holterman *et al.* (2006, 2011).

3. PCR (arthropods and nematodes)

Hot-start PCR master mixes facilitate PCR preparation as they enable the work to be done at room temperature and reduce possible errors. Great care has to be taken to avoid cross- and carry-over contamination during PCR preparation. Cross-contamination occurs when DNA from one sample gets into the PCR reaction of another sample. This mostly happens via contaminating primers, so whenever possible, PCR reactions should be prepared using sterile benches and laboratory rules should be strictly followed. The use of filter tips for all pipetting steps during PCR preparation and for aliquoting primers to avoid frequent use of the same tube over extended periods further helps to decrease the risk of contamination. Detecting cross-contamination can be done straightforwardly by including negative controls in every PCR run. If cross-contamination occurs despite the measures stated above, methodological alternatives, e.g. by using a modified nucleotide (dUTP, deoxyuridine triphosphate) and uracil-DNA N-glycosylase, can help. Carry-over contamination, the contamination of a PCR with amplicons from a previous PCR, also represents a serious threat. The general recommendation is to use different rooms (or well -separated compartments) for PCR preparation and for the sequencing reaction set-up. If a contamination event occurs, a complete decontamination of all surfaces, instruments and material is required. This may be done by UV irradiation and the use of DNA-destroying chemicals (see, for example, Champlot *et al*., 2010). Furthermore, it is generally recommended to change all the chemistry used for the PCR.

4. Analysis of PCR products (arthropods and nematodes)

After PCR, it is recommended that a test be done to see whether the reaction was successful and produced an amplicon of the expected size. This is usually done by agarose electrophoresis or by using automated electrophoresis equipment.

5. PCR product clean-up (arthropods and nematodes)

There are many kits and different technologies for rapid PCR product clean-up. Low-cost solutions such as ultrafiltration for quick and simple clean-up of PCR amplicons are sufficient.

6. DNA sequencing reaction (arthropods and nematodes)

Depending on the equipment used for in-house sequencing, the recommended chemistry (e.g. ABI dye terminator for ABI Sequencers from Applied Biosystems) should be used. Modifications to the manufacturer's recommendations, such as using

lower reaction volume and less sequencing mixture, should be thoroughly validated. Preparing two sequencing reactions per amplicon using the forward and the reverse PCR primers increases the rate of success for generating high-quality sequences.

7. Sequencing reaction clean-up (arthropods and nematodes)

There are many different products and technologies. Gel-filtration technology that elutes the DNA in a ready-to-sequence aqueous solution is one rapid and simple process that can be used.

8. DNA sequencing (ABI) (arthropods and nematodes)

The eluted sequencing reaction is placed into the sequencing instrument and manufacturer's protocols are followed to generate the DNA sequences.

9. Sequence data analysis (arthropods and nematodes)

The DNA sequences produced by Sanger sequencing (e.g. ABI sequencer from Applied Biosystems) are transferred into a software program that allows visualization of the chromatograms. The chromatograms are presented together with the sequencer's interpretation of each peak. The sequences can thus be checked for quality. The sequences may also be assembled, i.e. the two sequences produced with the forward and the reverse primer for each sample may be combined into one double-stranded product. The base calls of the software are then checked and corrected if necessary.

10. Species assignment (BLAST in GenBank; BOLD or Q-Bank) (arthropods and nematodes)

Based on its DNA sequence, each sample has then to be assigned to the corresponding species. This is only possible by comparing the query sequence with reference sequences contained in a database. The largest such database is GenBank, maintained by the US National Institutes of Health (NIH) (http://www.ncbi.nlm.nih.gov/genbank/). Alternatives that have the specific goal of enabling species identification using DNA barcoding are the Barcoding of Life Database (BOLD; http://www.boldsystems.org/) and the 'Comprehensive Databases on Quarantine Plant Pests and Diseases' (Q-Bank; http://www.q-bank.eu/), the first choice, especially for nematodes. These databases contain large numbers of reference sequences from many organisms with which the query sequence may be compared. The program generally used to perform this comparison is called Basic Local Alignment Search Tool or BLAST (www.ncbi.nlm.nih.gov/BLAST/). A search in these databases will result in a list of reference sequences with the closest match to the search sequence. In particular for nematodes, the Q-bank database allows for a multilocus BLAST search, in which the combination of up to three sequences (those for *LSU*, large-subunit ribosomal RNA gene; *SSU*, small-subunit ribosomal RNA gene; and *COI*, cytochrome c oxidase subunit I gene) can currently be used.

The output list of GenBank summarizes the results of the BLAST search, e.g. the title of the best matched database hits, the coverage of the overlapping fragment for both the query and the reference sequences, the best Expect value to indicate the probability of a random match and the highest percentage identity for the match (Max ident),

which indicates how close the two sequences are. Similar information is provided by Q-Bank which, in addition, provides a Rank score to assist in decision finding for species assignment. In both databases, it is up to the user to determine if the best hit is close enough to enable a species assignment of the query sequence. BOLD goes a step further in using a species gap or species differentiation threshold below which a query sequence is automatically assigned to the corresponding species.[1] If the BOLD search results in a species match, this is always confirmed by other evidence, such as BLAST searches in GenBank or on Q-Bank, combined with additional information such as accordance of geographic distribution and host plant.

Alternative approaches to aid in species assignment include phylogenetic tree building, character-based methods and statistical methods (see Goldstein and DeSalle, 2011, and references therein for a recent overview). Essentially, all methods are different grouping approaches that place the query sequence within or close to the most similar sequences in the database. Again, the problem for species assignment is how close is close enough to establish the assignment. For identification purposes, the ideal situation is where Max ident (GenBank) or the Similarity % (Q-Bank) of at least one database entry is 100%, indicating that the sequences are identical, which supports positive identification. Similarly, in the case of the Max ident hit list or the Similarity % of the Q-Bank, if the pairwise result list is close to 100% for one or several identical species and the corresponding value for the next species in the list is distinctly lower, e.g. by 5% or more, this may also be taken as evidence for a species assignment. However, there is general agreement that all available information should be used to assist in species assignment, i.e. an integrative approach (e.g. Padial *et al*., 2010, and references therein) should be used. The origin of the commodity and the specific host from which the specimen was collected support a species assignment. As the recommended databases contain different entries, at least in part, it is recommended that all three be used to provide an identification. If the reference sample of the pest species under consideration is contained in at least one of them, this should result in a reliable assignment to the species level, or at least to a highly similar species group, such as the *Bactrocera dorsalis* group, which arguably contains up to seven described species. As indicated above, the *COI* barcoding fragment is not able to differentiate among all arthropod species and so some species groups need to be accepted. If the main aim is qualifying a commodity as infested with quarantine pests originating from outside Europe, such species groups generally do not have to be further differentiated. Otherwise, the same procedures are used with an alternative gene, preferably with the highly variable ribosomal ITS regions 1 or 2.

Similar restrictions apply to nematodes as to arthropods, with a combination of two barcode sequences recommended for correct species assignment. With this multilocus identification approach, most species can be identified, with the exception of the tropical *Meloidogyne* species *M. incognita*, *M. arenaria* and *M. javanica*, although for these there are species-specific PCR assays that are able to identify these widespread species easily (Kiewnick *et al*., 2013).

Species assignment can be problematic if there are multiple species with a close to 100% match to the query sequence. This may be due to false entries in the databases or to specimens that were misidentified. If the outlier species occurs as a single entry among a larger number of another species, the hypothesis of it being a misidentified specimen may be justified and the query sample may be assigned to the other species. Such assignments should be communicated as being putative and efforts are needed to generate additional information, e.g. by sequencing another target gene and/or by carefully checking the geographic overlap of the assigned species with the origin of the commodity. Multiple species hits can also be caused by a lack of differentiation among closely related species, such as tephritids. In these cases, additional genetic markers for proper species assignment must be used. In some groups though, more sequence information may still not lead to differentiation, which indicates that the species status of the individual members of this group should be reconsidered by taxonomists. If species pairs cannot be differentiated, mitochondrial introgression caused by inter-species hybridization might have occurred. Under these circumstances, differences in nuclear genes are likely to exist, but as the two species share the same mitochondria after having hybridized, they cannot be differentiated with mitochondrial markers.

Note

1. The concept of a species threshold is still under debate, and has been shown to be inadequate in a number of animal species, e.g. van Velzen *et al*., 2012.

References

Acs, Z., Challis, R.J., Bihari, P., Blaxter, M., Hayward, A., Melika, G., Csoka, G., Penzes, Z., Pujade-Villar, J., Nieves-Aldrey, J.L. *et al*. (2010) Phylogeny and DNA barcoding of inquiline oak gallwasps (Hymenoptera: Cynipidae) of the western Palaearctic. *Molecular Phylogenetics and Evolution* 55, 210–225.

Ahrens, D., Monaghan, M.T. and Vogler, A.P. (2007) DNA-based taxonomy for associating adults and larvae in multi-species assemblages of chafers (Coleoptera: Scarabaeidae). *Molecular Phylogenetics and Evolution* 44, 436–449.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *Journal of Molecular Biology* 215, 403–410.

Arnot, D.E., Roper, C. and Bayoumi, R.A.L. (1993) Digital codes from hypervariable tandemly repeated DNA sequences in the *Plasmodium falciparum* circumsporozoite gene can genetically barcode isolates. *Molecular and Biochemical Parasitology* 61, 15–24.

Bhadury, P., Austen, M., Bilton, D., Lambshead, P., Rogers, A. and Smerdon, G. (2006) Development and evaluation of a DNA-barcoding approach for the rapid identification of nematodes. *Marine Ecology Progress Series* 320, 1–9.

Blaxter, M.L. (2004) The promise of DNA taxonomy. *Philosophical Transactions of the Royal Society B: Biological Sciences* 359, 669–679.

Blaxter, M.L., De Ley, P., Garey, J.R., Liu, L.X., Scheldeman, P., Vierstraete, A., Vanfleteren, J.R., Mackey, L.Y., Dorris, M., Frisse, L.M. *et al*. (1998) A molecular evolutionary framework for the phylum Nematoda. *Nature* 392, 71–75.

Blok, V.C., Phillips, M.S. and Harrower, B.E. (1997) Comparison of British populations of *Globoclera pallida* with populations from continental Europe and South America using RAPDs. *Genome* 40, 286–293.

Bonants, P., Groenewald, E., Rasplus, J.Y., Maes, M., De Vos, P., Frey, J.E., Boonham, N., Nicolaisen, M., Bertacini, A., Robert, V. *et al*. (2010) QBOL: a new EU project focusing on DNA barcoding of quarantine organisms. *EPPO Bulletin* 40, 30–33.

Boyce, T.M., Zwick, M.E. and Aquadro, C.F. (1989) Mitochondrial DNA in the bark weevils: size, structure and heteroplasmy. *Genetics* 123, 825–836.

Brunner, P.C. and Frey, J.E. (2010) Habitat-specific population structure in native western flower thrips *Frankliniella occidentalis* (Insecta, Thysanoptera). *Journal of Evolutionary Biology* 23, 797–804.

Brunner, P.C., Fleming, C. and Frey, J.E. (2002) A molecular identification key for economically important thrips species (Thysanoptera: Thripidae) using direct sequencing and a PCR-RFLP-based approach. *Agricultural and Forest Entomology* 4, 127–136.

Brunner, P.C., Chatzivassiliou, E.K., Katis, N.I. and Frey, J.E. (2004) Host-associated genetic differentiation in *Thrips tabaci* (Insecta; Thysanoptera), as determined from mtDNA sequence data. *Heredity* 93, 364–370.

Champlot, S., Berthelot, C., Pruvost, M., Bennett, E.A., Grange, T. and Geigl, E.-M. (2010) An efficient multistrategy DNA decontamination procedure of PCR reagents for hypersensitive PCR applications. *PLoS ONE* 5(9), e13042.

Chen, R., Jiang, L.Y. and Qiao, G.X. (2012) The effectiveness of three regions in mitochondrial genome for aphid DNA barcoding: a case in Lachininae. *PLoS ONE* 7(10), e46190.

Damm, S., Schierwater, B. and Hadrys, H. (2010) An integrative approach to species discovery in odonates: from character-based DNA barcoding to ecology. *Molecular Ecology* 19, 3881–3893.

De Ley, P., Tandingan De Ley, I., Morris, K., Abebe, E., Mundo-Ocampo, M., Yoder, M., Heras, J., Waumann, D., Rocha-Olivares, A., Burr, A.H.J. *et al*. (2005) An integrated approach to fast and informative morphological vouchering of nematodes for applications in molecular barcoding. *Philosophical Transactions of the Royal Society B: Biological Sciences* 360, 1945–1958.

De Luca, F., Vovlas, N., Lucarelli, G., Troccoli, A., Radicci, V., Fanelli, E., Cantalapiedra-Navarrete, C., Palomares-Rius, J.E. and Castillo, P. (2013) *Heterodera elachista* the Japanese

cyst nematode parasitizing corn in Northern Italy: integrative diagnosis and bionomics. *European Journal of Plant Pathology* 136, 857–872.

Derycke, S., Remerie, T., Vierstraete, A., Backeljau, T., Vanfleteren, J., Vincx, M. and Moens, T. (2005) Mitochondrial DNA variation and cryptic speciation within the free-living marine nematode *Pellioditis marina*. *Marine Ecology Progress Series* 300, 91–103.

Derycke, S., Backeljau, T., Vlaeminck, C., Vierstraete, A., Vanfleteren, J., Vincx, M. and Moens, T. (2006) Seasonal dynamics of population genetic structure in cryptic taxa of the *Pellioditis marina* complex (Nematoda : Rhabditida). *Genetica* 128, 307–321.

Derycke, S., Backeljau, T., Vlaeminck, C., Vierstraete, A., Vanfleteren, J., Vincx, M. and Moens, T. (2007) Spatiotemporal analysis of population genetic structure in *Geomonhystera disjuncta* (Nematoda, Monhysteridae) reveals high levels of molecular diversity. *Marine Biology* 151, 1799–1812.

Derycke, S., Vanaverbeke, J., Rigeaux, A., Backeljau, T. and Moens, T. (2010) Exploring the use of the cytochrome oxidase subunit 1 (COI) for DNA barcoding of free-living marine nematodes. *PLoS ONE* 5(10), e13716.

DeSalle, R., Egan, M.G. and Siddall, M. (2005) The unholy trinity: taxonomy, species delimitation and DNA barcoding. *Philosophical Transactions of the Royal Society B: Biological Sciences* 360, 1905–1916.

deWaard, J.R., Mitchell, A., Keena, M.A., Gopurenko, D., Boykin, L.M., Armstrong, K.F., Pogue, M.G., Lima, J., Floyd, R., Hanner, R.H. and Humble, L.M. (2010) Towards a global barcode library for *Lymantria* (Lepidoptera: Lymantriinae) tussock moths of biosecurity concern. *PLoS ONE* 5(12), e14280.

Dupuis, J.R., Roe, A.D. and Sperling, F.A.H. (2012) Multi-locus species delimitation in closely related animals and fungi: one marker is not enough. *Molecular Ecology* 21, 4422–4436.

Elias, M., Hill, R.I., Willmott, K.R., Dasmahapatra, K.K., Brower, A.V., Mallet, J. and Jiggins, C.D. (2007) Limited performance of DNA barcoding in a diverse community of tropical butterflies. *Proceedings of the Royal Society B: Biological Sciences* 274, 2881–2889.

Espineira, M., Herrero, B., Vieites, J.M. and Santaclara, F.J. (2010) Detection and identification of anisakids in seafood by fragment length polymorphism analysis and PCR-RFLP of ITS-1 region. *Food Control* 21, 1051–1060.

Evans, N. and Paulay, G. (2012) DNA barcoding methods for invertebrates. *Methods in Molecular Biology* 858, 47–77.

Floyd, R., Abebe, E., Papert, A. and Blaxter, M. (2002) Molecular barcodes for soil nematode identification. *Molecular Ecology* 11, 839–850.

Floyd, R.M., Rogers, A.D., Lambshead, P.J.D. and Smith, C.R. (2005) Nematode-specific PCR primers for the 18S small subunit rRNA gene. *Molecular Ecology Notes* 5, 611–612.

Floyd, R., Lima, J., deWaard, J., Humble, L. and Hanner, R. (2010) Common goals: policy implications of DNA barcoding as a protocol for identification of arthropod pests. *Biological Invasions* 12, 2947–2954.

Foley, D.H., Wilkerson, R.C., Cooper, R.D., Volovsek, M.E. and Bryan, J.H. (2007) A molecular phylogeny of *Anopheles annulipes* (Diptera: Culicidae) *sensu lato*: the most species-rich anopheline complex. *Molecular Phylogenetics and Evolution* 43, 283–297.

Folmer, O., Black, M., Hoeh, W., Lutz, R. and Vrijenhoek, R. (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology* 3, 294–299.

Frey, J.E. and Frey, B. (2004) Origin of intra-individual variation in PCR-amplified mitochondrial cytochrome oxidase I of *Thrips tabaci* (Thysanoptera: Thripidae): mitochondrial heteroplasmy or nuclear integration? *Hereditas* 140, 92–98.

Frey, J.E. and Pfunder, M. (2006) Molecular techniques for identification of quarantine insects and mites: the potential of microarrays. In: Rao, J.R., Fleming, C.C. and Moore, J.E. (eds) *Molecular Diagnostics: Current Technology and Applications*. Horizon Scientific, Norwich, UK, pp. 141–162.

Frey, J.E., Frey, B. and Baur, R. (2004) Molecular identification of the swede midge (Diptera: Cecidomyiidae). *The Canadian Entomologist* 136, 771-780.

Frey, J.E., Guillén, L., Frey, B., Samietz, J., Rull, J. and Aluja, M. (2013) Developing diagnostic SNP panels for the identification of true fruit flies (Diptera: Tephritidae) within the limits of COI-based species delimitation. *BMC Evolutionary Biology* 13, 106.

Goldstein, P.Z. and DeSalle, R. (2011) Integrating DNA barcode data and taxonomic practice: determination, discovery, and description. *BioEssays* 33, 135–147.

Gutiérrez-Gutiérrez, C., Cantalapiedra-Navarrete, C., Decraemer, W., Vovlas, N., Prior, T., Rius, J.E.P., and Castillo, P. (2012) Phylogeny, diversity, and species delimitation in some species of the *Xiphinema americanum*-group complex (Nematoda: Longidoridae), as inferred from nuclear and mitochondrial DNA sequences and morphology. *European Journal of Plant Pathology* 134, 561–597.

Hajibabaei, M., deWaard, J.R., Ivanova, N.V., Ratnasingham, S., Dooh, R.T., Kirk, S.L., Mackie, P.M. and Hebert, P.D.N. (2005) Critical factors for assembling a high volume of DNA barcodes. *Philosophical Transactions of the Royal Society B: Biological Sciences* 360, 1959–1967.

Hajibabaei, M., Janzen, D.H., Burns, J.M., Hallwachs, W. and Hebert, P.D.N. (2006) DNA barcodes distinguish species of tropical Lepidoptera. *Proceedings of the National Academy of Sciences of the United States of America* 103, 968–971.

Hajibabaei, M., Singer, G.A.C., Hebert, P.D.N. and Hickey, D.A. (2007) DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Trends in Genetics* 23, 167–172.

Hartig, G., Peters, R.S., Borner, J., Etzbauer, C., Misof, B. and Niehuis, O. (2012) Oligonucleotide primers for targeted amplification of single-copy nuclear genes in apocritan Hymenoptera. *PLoS ONE* 7(6), e39826.

Hawlitschek, O., Porch, N., Hendrich, L. and Balke, M. (2011) Ecological niche modelling and nDNA sequencing support a new, morphologically cryptic beetle species unveiled by DNA barcoding. *PLoS ONE* 6(2), e16662.

Hebert, P.D.N., Ratnasingham, S. and deWaard, J.R. (2003a) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society B: Biological Sciences* 270, 96-99.

Hebert, P.D., Cywinska, A., Ball, S.L. and deWaard, J.R. (2003b) Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences* 270, 313–321.

Hebert, P.D., Penton, E.H., Burns, J.M., Janzen, D.H. and Hallwachs, W. (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America* 101, 14812–14817.

Hemmerter, S., Slapeta, J. and Beebe, N.W. (2009) Resolving genetic diversity in Australasian *Culex* mosquitoes: incongruence between the mitochondrial cytochrome c oxidase I and nuclear acetylcholine esterase 2. *Molecular Phylogenetics and Evolution* 50, 317–325.

Holterman, M., Van der Wurff, A., Van den Elsen, S., Van Megen, H., Bongers, T., Holovachov, O., Bakker, J. and Helder, J. (2006) Phylum-wide analysis of SSU rDNA reveals deep phylogenetic relationships among nematodes and accelerated evolution toward crown clades. *Molecular Biology and Evolution* 23, 1792–1800.

Holterman, M., Rybarczyk, K., Van den Elsen, S., Van Megen, H., Mooyman, P., Pena-Santiago, R., Bongers, T., Bakker, J. and Helder, J. (2008) A ribosomal DNA-based framework for the detection and quantification of stress-sensitive nematode families in terrestrial habitats. *Molecular Ecology Resources* 8, 23–34.

Holterman, M., Karssen, G., Van den Elsen, S., Van Megen, H., Bakker, J. and Helder, J. (2009) Small subunit rDNA-based phylogeny of the Tylenchida sheds light on relationships among some high-impact plant-parasitic nematodes and the evolution of plant feeding. *Phytopathology* 99, 227–235.

Holterman, M.H.M., Frey, J.E., Helder, H., Mooyman, P.J.W., Rybarczyk, K.D. and Kiewnick, S. (2011) Barcoding quarantine nematodes and their close relatives: an update on the QBOL-project. *Communications in Agricultural and Applied Biological Sciences* 76, 403–407.

Hu, M., Hoglund, J., Chilton, N.B., Zhu, X.Q. and Gasser, R.B. (2002) Mutation scanning analysis of mitochondria cytochrome c oxidase subunit 1 reveals limited gene flow among bovine lungworm subpopulations in Sweden. *Electrophoresis* 23, 3357–3363.

Janzen, D.H. and Hallwachs, W. (2011) Joining inventory by parataxonomists with DNA barcoding of a large complex tropical conserved wildland in northwestern Costa Rica. *PLoS ONE* 6(8), e18123.

Ji, Y., Ashton, L., Pedley, S.M., Edwards, D.P., Tang, Y., Nakamura, A., Kitching, R., Dolman, P.M., Woodcock, P., Edwards, F.A. *et al*. (2013) Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters* 16, 1245–1257.

Khamis, F.M., Masiga, D.K., Mohamed, S.A., Salifu, D., De Meyer, M. and Ekesi, S. (2012) Taxonomic identity of the invasive fruit fly pest, *Bactrocera invadens*: concordance in morphometry and DNA barcoding. *PLoS ONE* 7(9), e44862.

Kiewnick, S., Holterman, M., Helder, H. and Frey, J. (2011) QBOL – barcoding as a new tool for identification of quarantine nematodes and their close relatives. *Phytopathology* 101, S90.

Kiewnick S., Braun-Kiewnick, A. and Frey, J.E. (2013) Development and validation of a real-time PCR assay for the detection and identification of the root-knot nematode *Meloidogyne enterolobii*, a new EPPO A2 list quarantine organism. *Journal of Plant Diseases and Protection* 3, 142.

Kim, M.I., Wan, X., Kim, M.J., Jeong, H.C., Ahn, N.H., Kim, K.G., Han, Y.S. and Kim, I. (2010) Phylogenetic relationships of true butterflies (Lepidoptera: Papilionoidea) inferred from COI, 16S rRNA and EF-1alpha sequences. *Molecules and Cells* 30, 409–425.

Landa, B.B., Rius, J.E.P., Vovlas, N., Carneiro, R., Maleita, C.M.N., Abrantes, I.M.D. and Castillo, P. (2008) Molecular characterization of *Meloidogyne hispanica* (Nematoda, Meloidogynidae) by phylogenetic analysis of genes within the rDNA in *Meloidogyne* spp. *Plant Disease* 92, 1104–1110.

Lee, W., Kim, H., Lim, J., Choi, H.R., Kim, Y., Kim, Y.S., Ji, J.Y., Foottit, R.G. and Lee, S. (2011) Barcoding aphids (Hemiptera: Aphididae) of the Korean Peninsula: updating the global data set. *Molecular Ecology Resources* 11, 32–37.

Lefebure, T., Douady, C.J., Gouy, M. and Gibert, J. (2006) Relationship between morphological taxonomy and molecular divergence within Crustacea: proposal of a molecular threshold to help species delimitation. *Molecular Phylogenetics and Evolution* 40, 435–447.

Leo, S.S., Pybus, M.J. and Sperling, F.A. (2010) Deep mitochondrial DNA lineage divergences within Alberta populations of *Dermacentor albipictus* (Acari: Ixodidae) do not indicate distinct species. *Journal of Medical Entomology* 47, 565–574.

Lipscomb, D., Platnick, N. and Wheeler, Q. (2003) The intellectual content of taxonomy: a comment on DNA taxonomy. *Trends in Ecology and Evolution* 18, 65–66.

Madani, M., Subbotin, S.A., Ward, L.J., Li, X. and De Boer, S.H. (2010) Molecular characterization of Canadian populations of potato cyst nematodes, *Globodera rostochiensis* and *G. pallida* using ribosomal nuclear RNA and cytochrome b genes. *Canadian Journal of Plant Pathology* 32, 252–263.

Magnacca, K.N. and Brown, M.J. (2010) Mitochondrial heteroplasmy and DNA barcoding in Hawaiian *Hylaeus* (*Nesoprosopis*) bees (Hymenoptera: Colletidae). *BMC Evolutionary Biology* 10,174.

Maneesakorn, P., An, R.S., Daneshvar, H., Taylor, K., Bai, X.D., Adams, B.J., Grewal, P.S. and Chandrapatya, A. (2011) Phylogenetic and cophylogenetic relationships of entomopathogenic nematodes (*Heterorhabditis*: Rhabditida) and their symbiotic bacteria (*Photorhabdus*: Enterobacteriaceae). *Molecular Phylogenetics and Evolution* 59, 271–280.

Marek, M., Zouhar, M., Douda, O., Mazakova, J. and Rysanek, P. (2010) Bioinformatics-assisted characterization of the ITS1-58.S-ITS2 segments of nuclear rRNA gene clusters, and its

exploitation in molecular diagnostics of European crop-parasitic nematodes of the genus *Ditylenchus*. *Plant Pathology* 59, 931–943.

Matz, M.V. and Nielsen, R. (2005) A likelihood ratio test for species membership based on DNA sequence data. *Philosophical Transactions of the Royal Society B: Biological Sciences* 360, 1969–1974.

Meier, R., Shiyang, K., Vaidya, G. and Ng, P.K.L. (2006) DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Systematic Biology* 55, 715–728.

Meyer, C.P. and Paulay, G. (2005) DNA barcoding: error rates based on comprehensive sampling. *PLoS Biology* 3(12), e422.

Monaghan, M.T., Balke, M., Gregory, T.R. and Vogler, A.P. (2005) DNA-based species delineation in tropical beetles using mitochondrial and nuclear markers. *Philosophical Transactions of the Royal Society B: Biological Sciences* 360, 1925–1933.

Moulton, M.J., Song, H. and Whiting, M.F. (2010) Assessing the effects of primer specificity on eliminating numt coamplification in DNA barcoding: a case study from Orthoptera (Arthropoda: Insecta). *Molecular Ecology Resources* 10, 615–627.

Nagoshi, R.N., Brambila, J. and Meagher, R.L. (2011) Use of DNA barcodes to identify invasive armyworm *Spodoptera* species in Florida. *Journal of Insect Science* 11, 154.

Nielsen, R. and Matz, M. (2006) Statistical approaches for DNA barcoding. *Systematic Biology* 55, 162–169.

Odegaard, F. (2000) How many species of arthropods? Erwin's estimate revised. *Biological Journal of the Linnean Society* 71, 583–597.

Padial, J.M., Miralles, A., De la Riva, I. and Vences, M. (2010) The integrative future of taxonomy. *Frontiers in Zoology* 7, 16.

Pamilo, P., Viljakainen, L. and Vihavainen, A. (2007) Exceptionally high density of NUMTs in the honeybee genome. *Molecular Biology and Evolution* 24, 1340–1346.

Park, D.S., Suh, S.J., Oh, H.W. and Hebert, P.D.N. (2010) Recovery of the mitochondrial COI barcode region in diverse Hexapoda through tRNA-based primers. *BMC Genomics* 11, 423.

Park, D.S., Foottit, R., Maw, E. and Hebert, P.D.N. (2011a) Barcoding bugs: DNA-based identification of the true bugs (Insecta: Hemiptera: Heteroptera). *PLoS ONE* 6(4), e18749.

Park, D.S., Suh, S.J., Hebert, P.D.N., Oh, H.W. and Hong, K.J. (2011b) DNA barcodes for two scale insect families, mealybugs (Hemiptera: Pseudococcidae) and armored scales (Hemiptera: Diaspididae). *Bulletin of Entomological Research* 101, 429–434.

Porazinska, D.L., Giblin-Davis, R.M., Faller, L., Farmerie, W., Kanzaki, N., Morris, K., Powers, T.O., Tucker, A.E., Sung, W. and Thomas, W. K. (2009) Evaluating high-throughput sequencing as a method for metagenomic analysis of nematode diversity. *Molecular Ecology Resources* 9, 1439–1450.

Powers, T. (2004) Nematode molecular diagnostics: from bands to barcodes. *Annual Review of Phytopathology* 42, 367–383.

Powers, T., Harris, T., Higgins, R., Mullin, P., Sutton, L. and Powers, K. (2011) MOTUs, morphology, and biodiversity estimation: a case study using nematodes of the suborder Criconematina and a conserved 18S DNA barcode. *Journal of Nematology* 43, 35–48.

Rach, J., DeSalle, R., Sarkar, I.N., Schierwater, B. and Hadrys, H. (2008) Character-based DNA barcoding allows discrimination of genera, species and populations in Odonata. *Proceedings of the Royal Society B: Biological Sciences* 275, 237–247.

Ratnasingham, S. and Hebert, P.D.N. (2007) BOLD: the barcode of life data system (http://www.barcodinglife.org). *Molecular Ecology Notes* 7, 355–364.

Rebijith, K.B., Asokan, R., Kumar, N.K.K., Srikumar, K.K., Ramamurthy, V.V. and Bhat, P.S. (2012) DNA barcoding and development of species-specific markers for the identification of tea mosquito bugs (Miridae: Heteroptera) in India. *Environmental Entomology* 41, 1239–1245.

Remigio, E.A. and Hebert, P.D.N. (2003) Testing the utility of partial COI sequences for phylogenetic estimates of gastropod relationships. *Molecular Phylogenetics and Evolution* 29, 641–647.

Ros, V.D. and Breeuwer, J.J. (2007) Spider mite (Acari: Tetranychidae) mitochondrial COI phylogeny reviewed: host plant relationships, phylogeography, reproductive parasites and barcoding. *Experimental and Applied Acarology* 42, 239–262.

Rubinoff, D. and Holland, B.S. (2005) Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. *Systematic Biology* 54, 952–961.

Rubinoff, D., Cameron, S. and Will, K. (2006) A genomic perspective on the shortcomings of mitochondrial DNA for "Barcoding" identification. *Journal of Heredity* 97, 581–594.

Rybarczyk-Mydłowska, K., Mooyman, P., Van Megen, H., Van den Elsen, S., Vervoort, M., Veenhuizen, P., Van Doom, J., Dees, R., Karssen, G., Bakker, J. and Helder, J. (2012) Small subunit ribosomal DNA-based phylogenetic analysis of foliar nematodes (*Aphelenchoides* spp.) and their quantitative detection in complex DNA backgrounds. *Phytopathology* 102, 1153–1160.

Saitou, N. and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4, 406–425.

Sarkar, I.N., Planet, P.J. and DeSalle, R.O.B. (2008) caos software for use in character-based DNA barcoding. *Molecular Ecology Resources* 8, 1256–1259.

Saunders, G.W. (2005) Applying DNA barcoding to red macroalgae: a preliminary appraisal holds promise for future applications. *Philosophical Transactions of the Royal Society B: Biological Sciences* 360, 1879–1888.

Seifert, K.A. (2009) Progress towards DNA barcoding of fungi. *Molecular Ecology Resources* 9, 83–89.

Shatters, R.G., Jr, Powell, C.A., Boykin, L.M., Liansheng, H. and Mckenzie, C.L. (2009) Improved DNA barcoding method for *Bemisia tabaci* and related Aleyrodidae: development of universal and *Bemisia tabaci* biotype-specific mitochondrial cytochrome c oxidase I polymerase chain reaction primers. *Journal of Economic Entomology* 102, 750–758.

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H. and Flook, P. (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene-sequences and a compilation of conserved polymerase chain-reaction primers. *Annals of the Entomological Society of America* 87, 651–701.

Simon, C., Buckley, T.R., Frati, F., Stewart, J.B. and Beckenbach, A.T. (2006) Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved polymerase chain reaction primers for animal mitochondrial DNA. *Annual Review of Ecology Evolution and Systematics* 37, 545–579.

Smith, C.I., Drummond, C.S., Godsoe, W., Yoder, J.B. and Pellmyr, O. (2009) Host specificity and reproductive success of yucca moths (*Tegeticula* spp. Lepidoptera: Prodoxidae) mirror patterns of gene flow between host plant varieties of the Joshua tree (*Yucca brevifolia*: Agavaceae). *Molecular Ecology* 18, 5218–5229.

Song, H., Buhay, J.E., Whiting, M.F. and Crandall, K.A. (2008) Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proceedings of the National Academy of Sciences of the United States of America* 105, 13486–13491.

Subbotin, S.A., Halford, P.D., Warry, A. and Perry, R.N. (2000) Variations in ribosomal DNA sequences and phylogeny of *Globodera* parasitising solanaceous plants. *Nematology* 2, 591–604.

Subbotin, S.A., Vovlas, N., Crozzoli, R., Sturhan, D., Lamberti, F., Moens, M. and Baldwin, J.G. (2005) Phylogeny of Criconematina Siddiqi, 1980 (Nematoda: Tylenchida) based on morphology and D2–D3 expansion segments of the 28S-rRNA gene sequences with application of a secondary structure model. *Nematology* 7, 927–944.

Sutou, M., Kato, T. and Ito, M. (2011) Recent discoveries of armyworms in Japan and their species identification using DNA barcoding. *Molecular Ecology Resources* 11, 992–1001.

Tandingan De Ley, I., De Ley, P., Vierstraete, A., Karssen, G., Moens, M. and Vanfleteren, J. (2002) Phylogenetic analyses of *Meloidogyne* small subunit rDNA. *Journal of Nematology* 34, 319–327.

Tanzler, R., Sagata, K., Surbakti, S., Balke, M. and Riedel, A. (2012) DNA barcoding for community ecology – how to tackle a hyperdiverse, mostly undescribed Melanesian fauna. *PLoS ONE* 7(1), e28832.

Taylor, H.R. and Harris, W.E. (2012) An emergent science on the brink of irrelevance: a review of the past 8 years of DNA barcoding. *Molecular Ecology Resources* 12, 377–388.

Thormann, B., Raupach, M.J., Wagner, T., Wagele, J.W. and Peters, M.K. (2011) Testing a short nuclear marker for inferring staphylinid beetle diversity in an African tropical rain forest. *PLoS ONE* 6(3), e18101.

Timm, A.E., Stiller, M. and Frey, J.E. (2008) A molecular identification key for economically important thrips species (Thysanoptera: Thripidae) in southern Africa. *African Entomology* 16, 68–75.

Toumi, F., Waeyenberge, L., Viaene, N., Dababat, A., Nicol, J.M., Ogbonnaya, F. and Moens, M. (2013) Development of two species-specific primer sets to detect the cereal cyst nematodes *Heterodera avenae* and *Heterodera filipjevi*. *European Journal of Plant Pathology* 136, 613–624.

Valentini, A., Pompanon, F. and Taberlet, P. (2009) DNA barcoding for ecologists. *Trends in Ecology and Evolution* 24, 110–117.

Van Megen, H., Van den Elsen, S., Holterman, M., Karssen, G., Mooyman, P., Bongers, T., Holovachov, O., Bakker, J. and Helder, J. (2009) A phylogenetic tree of nematodes based on about 1200 full-length small subunit ribosomal DNA sequences. *Nematology* 11, 927–950.

van Velzen, R., Weitschek, E., Felici, G. and Bakker, F.T. (2012) DNA barcoding of recently diverged species: relative performance of matching methods. *PLoS ONE* 7(1), e30490.

Vences, M., Thomas, M., Bonett, R.M. and Vieites, D.R. (2005) Deciphering amphibian diversity through DNA barcoding: chances and challenges. *Philosophical Transactions of the Royal Society B: Biological Sciences* 360, 1859–1868.

Vovlas, N., Trisciuzzi, N., Troccoli, A., Luca, F., Cantalapiedra-Navarrete, C. and Castillo, P. (2013) Integrative diagnosis and parasitic habits of *Cryphodera brinkmani* a non-cyst forming heteroderid nematode intercepted on Japanese white pine bonsai trees imported into Italy. *European Journal of Plant Pathology* 135, 717–726.

Waugh, J. (2007) DNA barcoding in animal species: progress, potential and pitfalls. *BioEssays* 29, 188–197.

Will, K.W., Mishler, B.D. and Wheeler, Q.D. (2005) The perils of DNA barcoding and the need for integrative taxonomy. *Systematic Biology* 54, 844–851.

Wilson, J.J. (2012) DNA barcodes for insects. *Methods in Molecular Biology* 858, 17–46.

Yang, Q.Q., Zhao, S., Kucerova, Z., Stejskal, V., Opit, G., Qin, M., Cao, Y., Li, F.J. and Li, Z.H. (2013) Validation of the 16S rDNA and COI DNA barcoding technique for rapid molecular identification of stored product psocids (Insecta: Psocodea: Liposcelididae). *Journal of Economic Entomology* 106, 419–425.

Ye, W., Giblin-Davis, R.M., Davies, K.A., Purcell, M.F., Scheffer, S.J., Taylor, G.S., Center, T.D., Morris, K. and Thomas, W.K. (2007) Molecular phylogenetics and the evolution of host plant associations in the nematode genus *Fergusobia* (Tylenchida: Fergusobiinae). *Molecular Phylogenetics and Evolution* 45, 123–141.

Zhang, D.X. and Hewitt, G.M. (1996) Nuclear integrations: challenges for mitochondrial DNA markers. *Trends in Ecology and Evolution* 11, 247–251.

Zijlstra, C., Lever, A.E.M., Uenk, B.J. and Van Sil-Fhout, C.H. (1995) Differences between ITS regions of isolates of the root-knot nematodes *Meloidogyne hapla* and *M. chitwoodi*. *Phytopathology* 85, 1231–1237.

7 Microarrays

Ian Adams,* Jenny Tomlinson and Catherine Harrison

Fera, Sand Hutton, York, UK

7.1 Introduction

DNA microarrays consist of multiple probes attached to a solid surface for the parallel detection of multiple targets. Nucleic acid from the sample is hybridized to the arrayed probes, and by examining which probes have hybridized with the sample, it is possible to identify the sequences present within the sample (Fig. 7.1). The use of microarrays in plant disease diagnosis allows tens or thousands of individual tests to be carried out in a single assay. For plant virus testing, microarrays are usually more sensitive than ELISA but less sensitive than real-time PCR assays (Boonham *et al.*, 2003); for the routine testing of a defined set of pathogens, microarrays are cheaper than next-generation sequencing.

Microarrays were developed in 1995 as an extension of Southern blot-based membrane arrays and were first used to study gene expression (Schena *et al.*, 1995). These first arrays consisted of tens or hundreds of cloned PCR products printed on to glass microscope slides, which were hybridized with fluorescently labelled sample cDNA. The production of arrays of PCR products is time-consuming, and the length of the products used can have a significant effect on specificity. A significant development was the use of pre-synthesized oligonucleotides as probes, which increased consistency and allowed the density of printed arrays to increase to a maximum of about 10,000 probes. The use of shorter oligonucleotides (typically 20–70 bases) also allows probes to be designed using available sequence information to manipulate properties such as sensitivity and specificity. The incorporation of modifications (typically an amino group) at the 3′ or 5′ terminus allows the probe to be attached to the functionalized

*Corresponding author: ian.adams@fera.co.uk

glass surface in a specific orientation, which can also affect hybridization to the target (Boonham *et al.*, 2007). Figure 7.2 shows a typical glass slide-based oligonucleotide array hybridized with fluorescently labelled cDNA.

The next major advance in array manufacture was the use of *in situ* synthesis of oligonucleotide probes directly on the array surface. This technology allows the probe density to be increased to over 2 million per array (as, for example, in the Affymetrix custom GW chip arrays). As well as the extremely high probe density, on-chip synthesis has the advantage of flexibility as probe sequences can easily be added or removed. However, the cost per array is relatively high, and the high probe density is not required for many diagnostic applications.

In order to identify the probes to which the sample DNA has bound, it is necessary to label the sample DNA before hybridization to the array. A common approach is to use fluorescent dyes, which may be incorporated directly using labelled primers or nucleotides for PCR or reverse transcription. An alternative approach is the use of nucleotides with other modifications, which may be more efficiently incorporated, such as aminoallyl groups which can then be labelled prior to hybridization using reactive esters of fluorescent dyes such as Cy3 and Cy5 (Boonham *et al.*, 2007). A commonly used experimental design is to hybridize two samples to the same array, e.g. when comparing the nucleic acid from a healthy plant labelled with one dye (typically Cy3) with that from

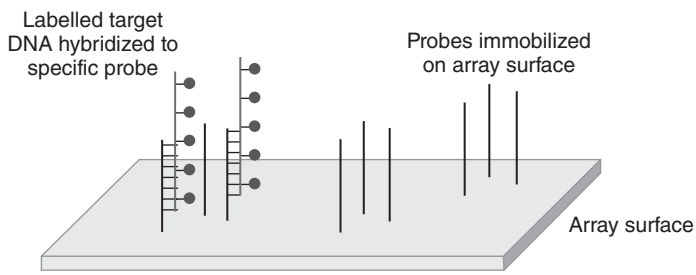


Fig. 7.1. Diagram showing microarray probes, the labelled target nucleic acid and the array surface.

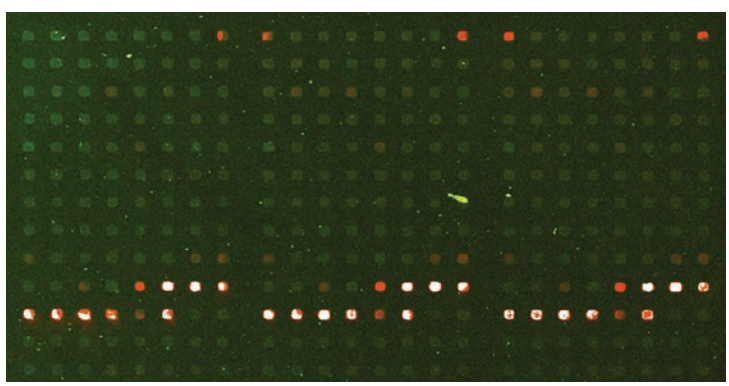


Fig. 7.2. An example glass slide array hybridized with cDNA from a plant infected with *Potato virus Y* labelled with the fluorescent dye Cy5. The array is shown printed in triplicate.

a diseased plant labelled with another dye (typically Cy5). Non-fluorescent detection is also possible, for example using antibody-based detection of incorporated labels such as biotin, as described in Section 7.3. Labelling may be achieved with or without the amplification of the sample nucleic acid, using specific or non-specific primers. RNA viruses can be detected or gene expression analysed by incorporating labels during reverse transcription using non-specific primers such as random hexamers or oligo-dT (or oligo(dT); a short sequence of deoxythymine nucleotides). For the detection of DNA targets, a common approach is to use PCR with specific primers. Some pathogens can be detected by using generic primers (e.g. targeting conserved 16S sequences for groups of bacteria), and different species can be resolved using target-specific probes. Alternatively, DNA from each target species may be amplified using specific primers in multiplex PCR.

Microarrays were initially used mostly for gene expression studies, but they have now been largely replaced for this purpose by next-generation sequencing (see Chapter 8). Arrays have been used extensively in cancer research for comparative genomics, the detection of single nucleotide polymorphisms (SNPs), copy number variation and fusion gene detection. They have also been used for pathogen detection, as already mentioned, and this will be the focus of this chapter.

7.2 Microarrays for pathogen detection

Microarrays were first used for the detection of viral pathogens in 2002 when Wang *et al.* (2002) developed a human virus array using oligonucleotide probes. A later version of this array was used in the identification of a novel coronavirus as the probable cause of severe acute respiratory syndrome (SARS) (Wang *et al.*, 2003). Plant virus arrays were first developed in 2003 when Boonham *et al.* (2003) and Bystricka *et al.* (2003) developed cDNA-based potato virus arrays, and Lee *et al.* (2003) developed a cucurbit virus array. In each case, the use of the arrays was demonstrated for species-level detection of viruses in single and/or mixed infections. The use of oligonucleotide arrays for plant virus detection was first described by Bystricka *et al.* (2005). Glass slide microarrays have since been applied to the detection of viruses of grapevine (Engel *et al.*, 2010), tomato (Tiberini and Barba, 2012) and other hosts (Nicolaisen, 2011), for *Plum pox virus* strain discrimination (Pasquini *et al.*, 2008) and in the detection of plant viruses and viroids at genus level (Zhang *et al.*, 2010, 2013).

Glass slide-based arrays showed great promise for plant virus detection, both as a routine diagnostic tool and for novel virus discovery, but their use in the latter role has now been replaced by next-generation sequencing (see Chapter 8). Despite their early promise and demonstrable potential, conventional glass slide microarrays have not been adopted for routine diagnostic use, primarily because of the complexity of the required workflows and a lack of robustness and repeatability.

7.3 Alternative microarray platforms

A number of microarray platforms have been developed specifically to enable disease diagnostics, and these have attempted to overcome the problems of complexity and robustness in different ways. CombiMatrix developed a silicon chip-based oligonucleotide

array platform mainly for the biodefence market. The chip contains circuits that allow the detection of DNA binding to the array electronically, thus simplifying the workflow, as no labelling or post-hybridization scanning is required (Ghindilis *et al.*, 2007).

An alternative approach was taken by Life Technologies with the Luminex xMAP® technology (from Luminex). This system uses 5.6 µm beads filled with one of 500 different dye mixtures and coated with specific capture oligonucleotides. Fluorescently labelled sample DNA is hybridized with the beads, and the bead mixture is passed through a modified flow cytometer; this determines the dye mixture of each bead, identifying which capture oligonucleotide is attached to a specific bead and detecting the presence of labelled sample DNA. Using this approach, a cocktail of different beads can be made to test for up to 500 different targets. This technology is currently being trialled in a number of UK hospital laboratories for the rapid detection of gastrointestinal pathogens (Pankhurst *et al.*, 2014), and has been applied to the detection of plant viruses and viroids (van Brunschot *et al.*, 2014a,b). Ten probe/bead combinations were developed and found to robustly detect the viroids tested. Furthermore, the sensitivity of the plant virus bead array was found to be equivalent to that of real-time PCR for at least one of the target viruses.

So-called 'lab-on-a-chip' methods are also being developed, which integrate simple microarrays with electronic circuitry, allowing automated detection of DNA hybridization with an on-chip PCR reaction chamber. This allows nucleic acid to be added to the chip, which then carries out the microarray assay and provides a read-out of the results. Such a device has been demonstrated as being capable of discriminating between plant pathogenic *Phytophthora* species (Julich *et al.*, 2011) and shows great promise, but is currently not ready for routine use.

Alere has taken a different approach to improving the robustness of glass slide microarrays. This company's 'Array Tubes' consist of 4 × 4 mm glass slide arrays mounted in the bottom of 1.5 ml centrifuge tubes and printed with up to 540 oligonucleotide features. Figure 7.3 shows an Alere array tube. PCR using biotinylated nucleotides

Fig. 7.3. The Alere Array Tube showing the glass slide array mounted in the base of a centrifuge tube. The oligonucleotides on the array can be seen as small spots in a square grid pattern. The larger square features on the array are used for orientation.

is used to label the sample DNA, which is hybridized to the array in the tube. The biotin is then detected using horseradish peroxidase (HRP)-linked streptavidin or anti-biotin antibody, and a colorimetric substrate allows visualization of the bound sample DNA using a simple scanner (Fig. 7.4). The simplicity of the tube-based workflow makes the assay repeatable and robust in comparison with arrays in a conventional glass slide format. Array tube assays have been developed for bacteria (Batchelor *et al.*, 2008; Cannon *et al.*, 2010; Tomlinson *et al.*, 2014), fungi (Monecke *et al.*, 2006) and human respiratory viruses (Cannon *et al.*, 2010), and array tubes for antibiotic resistance have also been successfully marketed as a service (Walsh *et al.*, 2010). An array has also been developed for the detection of 30 different potato viruses (Barrett *et al.*, 2009), and Fig. 7.5 shows this array hybridized with nucleic acid from a virus-infected potato.

Fig. 7.4. The Alere ATR03 scanner with Array Tube in place.

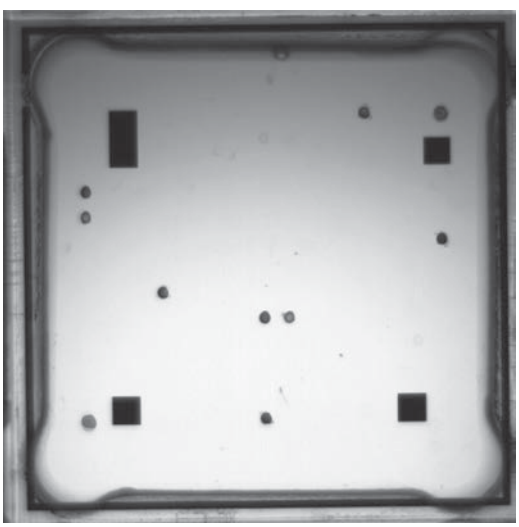


Fig. 7.5. Example slide array from an Alere Array Tube showing a potato array hybridized with a potato plant infected with *Potato virus M*.

7.4 Array design

The principles of oligonucleotide probe design are similar regardless of the particular array format. For virus detection, oligonucleotide probes are typically 50–70 nucleotides in length, with a melting temperature of 58–60°C. Many plant viruses have an RNA genome, and different isolates can vary by as much as 20% and still be considered part of the same species (King *et al.*, 2012). Longer probes with higher melting temperatures allow some mismatches to be tolerated so that this variation does not cause false negative results. Probes for the detection of bacteria, fungi and insects, where small sequence differences or even a single mismatch may indicate different species, need to be shorter (20–25 nucleotides), and typically with lower melting temperatures. By designing probes of different lengths to tolerate or discriminate sequence differences, it is possible to combine the detection of viruses, bacteria, fungi and insects on a single array.

Microarray probes are designed in a similar way to PCR primers. Nucleotide sequence alignments of target and related species are constructed to identify regions that are conserved within a species but different in other species. The melting temperature and secondary structure are then checked. It is common to design multiple oligonucleotides for each target species to increase the robustness of detection; for bacterial pathogens, a useful approach is to target both chromosomal and plasmid sequences in order to obtain the most relevant information about the disease status of the host plant.

Probes may be tested empirically and the best are then used for diagnostic arrays. Not all oligonucleotides that look good when designed perform well. Reasons for probe failure can include secondary structure in the target DNA (Peplies *et al.*, 2003; Wei *et al.*, 2012), or unexpected cross-reactivity with non-target sequences. It is usually necessary go through the cycle of design, synthesis, testing and analysis a number of times to produce an optimally functioning array. Controls should include melting temperature-matched negative control oligonucleotides that do not match any published sequences on GenBank (the US National Institutes of Health genetic sequence database available from NCBI, the US National Center for Biotechnology Information; see Benson *et al.*, 2010) and plant control probes to detect the host plant, for example, by targeting the cytochrome oxidase I or *RbcL* genes.

7.5 Conclusions

In 2008, an article in *Nature* suggested that the increasing popularity of next-generation sequencing might mean the 'death of microarrays' (Ledford, 2008). For many of the large markets for microarrays, such as gene expression studies and cancer research, this has in fact happened. In disease diagnostics, however, due to the development of diagnostic-centred microarray platforms such as tube and bead arrays, the use of microarrays is likely to continue to increase, as they allow relatively cheap routine testing for defined sets of pathogens. As discussed in Chapter 8, next-generation sequencing can be used for diagnostics, but the cost of sequencing and, particularly, the cost and difficulty of analysis is likely to keep this from routine use for all but the biggest laboratories for at least the next few years.

Protocols

Contents

Protocol 7.1

Glass slide array protocol for detection of RNA plant viruses

Materials

- Printed oligonucleotide arrays (see Note 1)
- Cy3- and/or Cy5-(fluorescent dye) labelled first strand cDNA or amplified DNA (see Note 2)
- Laminar flow cabinet for setting up amplification reactions (if required, see Step 1 below)
- Nexterion hybridization buffer (Schott UK)
- Gene Frames (Thermo Scientific)
- Hybridization chamber
- Incubator
- Heated block or PCR machine
- Wash buffer I (2 × SSC (saline-sodium citrate), 0.1% SDS (sodium dodecyl sulfate))
- Wash buffer II (1 × SSC)
- Wash buffer III (0.1 × SSC)
- Air duster
- Array scanner, e.g. GenePix 4000B (Molecular Devices)

Method

1. Prior to hybridization to an array, the sample RNA is converted into fluorescently labelled first strand cDNA. This can be achieved in different ways, with or without amplification. Non-amplification methods require large quantities of RNA (30 µg). Methods using the random amplification method require less input RNA (1–5 µg) but can be prone to amplification bias, leading to some viral sequences being over-represented. Each method is described below:

(i) Non-amplified Amersham CyScribe Post-Labeling Kit (GE Healthcare): with this kit, first strand cDNA is produced from the RNA by reverse transcription primed with random nonamers and/or oligo(dT) (a short sequence of deoxythymine nucleotides) primers, using aminoallyl-dUTP (deoxyuridine triphosphate) in place of dTTP (deoxythymidine triphosphate). In a second step, the aminoallyl-labelled DNA is reacted with florescent Cy5- or Cy3-NHS (*N*-hydroxysuccinimide) esters to produce Cy5- or Cy3-labelled DNA. This DNA is then purified and is ready for hybridization to the array. Detailed protocols are available from the manufacturer.

(ii) Randomly amplified labelled cDNA: this method is adapted from Wang *et al*. (2002) and uses random octamers, tagged with a binding site for a PCR primer (NNNNNNNNCGCCGTTTCCCAGTAGGTCTC), (see Note 3) to prime reverse transcription of the first DNA strand. The second strand is synthesized using a limiting amount of tagged random primer and an excess of tag primer (CGCCGTTTCCCAGTAGGTCTC). The fluorescent label is incorporated by using Cy5- or Cy3-modified dCTP (deoxycytidine triphosphate) in the PCR reaction.

2. Incubate labelled DNA in hybridization buffer at 95°C for 3 min, then place on ice for up to 3 min.

3. Apply a gene frame to the surface of the slide and remove the backing sheet. Apply the prepared sample DNA to the array inside the gene frame and carefully apply a polyester cover slip.

4. Place the prepared array into a hybridization chamber or sealed box with a small amount of water to create a humid environment. Hybridize overnight at 42°C in the dark.

5. Pre-warm wash buffers I, II and III. Place the array in a suitable container, for example, a slide mailer tube. Incubate the slide in wash buffer I for 5 min; discard the buffer and add wash buffer II for 5 min, then discard and add wash buffer III for 5 min.

6. Carefully dry the slide using compressed air or an aerosol air duster and store in a dark container until ready to scan using a fluorescent scanner such as the GenePix 4000B (Molecular Devices).

7. The GenePix 4000B scanner allows fluorescently labelled arrays to be scanned with excitation at 635 nm and 532 nm, corresponding to the fluorophores Cy5 and Cy3. In order to avoid saturated signals, it is possible to control both the power of the laser used for excitation and the gain of the photomultiplier tubes used for detection of the emitted light.

8. Spot intensity and background can be determined using the software supplied with the scanner (for example, GenePix Pro Microarray analysis software provided with the 4000B scanner). Averages can be taken across replicates of the same probes, and also of multiple probes, and the averages plotted against the average for the negative control probes. An ANOVA statistical test can be carried out with

a Dunnett's post hoc test to determine whether any set of probes has a signal significantly above that of the negative control. An alternative approach is to use the DetectiV software package (Watson *et al*., 2007), which allows visualization and normalization of array data. The package then uses a *t*-test to determine whether any of the probes has a normalized signal significantly above zero.

Notes

1. Conventional glass slides can be printed by various suppliers using a range of chemistries (functionalization of surface and probes). Probes are commonly represented more than once on the array to allow average signals to be accounted for.

2. Labelled DNA can be generated from sample RNA using a range of different methods, see Step 1 above. RNA can be extracted using any method that yields RT-PCR grade RNA. See Chapter 2 for a discussion of RNA extraction.

3. A, adenine; C, cytosine; G, guanine; T, thymine; N, unknown base.

Protocol 7.2

Array tube protocol

Materials

- Custom array tubes (Alere) (see Note 1)
- PCR reagents: *Taq* DNA polymerase, such as GoTaq® (Promega) and PCR buffer, $MgCl_2$ and dNTPs (deoxynucleotide triphosphates)
- Biotin dCTP (deoxycytidine triphosphate; Jena). Biotin dUTP (deoxyuridine triphosphate) can also be used
- Target-specific primers (see Note 2)
- Pipettors – separate sets required for setting up reactions (Steps 1–3), adding DNA to reactions (Step 4) and post-PCR (Steps 7–21).
- Filter tips for pipettors
- Vortexer
- 96 well plate or 0.2 ml reaction tubes
- PCR machine
- Laminar flow cabinet for setting up reactions
- QIAquick PCR Purification Kit (Qiagen)
- Nexterion hybridization buffer (Schott UK)
- Molecular biology grade water
- Wash buffer I (2 × SSC (saline-sodium citrate), 0.01% Triton X-100)
- Wash buffer II (2 × SSC)
- Wash buffer III (0.2 × SSC)
- Phosphate-buffered saline (PBS)
- Blocking solution (PBS, 0.02% milk powder, 0.002% Triton X-100)
- Conjugation solution (blocking solution, 1% antibody)

- HRP (horseradish peroxidase)-linked anti-biotin antibody (New England Biolabs) (see Note 3)
- PBS containing 0.1% Tween 20
- SeramunGrün chip substrate (Seramun Diagnostica)
- Thermal shaker or shaking incubator
- 0.6 ml and 2 ml tubes
- ATR03 scanner (Alere) and Iconoclust software

Method

1. Reactions should be set up in a laminar flow cabinet to prevent contamination; DNA and RNA should not be handled in the cabinet.
2. Set up the reactions containing the following components. Make up a master mix (see Protocol 7.2, Table 1) for the number of reactions to be run plus at least four extra reactions to allow for pipetting inaccuracies.
3. Vortex briefly, then pipette 24 µl aliquots of master mix into the wells of a 96 well plate or into 0.2 ml reaction tubes.
4. Add 1 µl sample DNA extract, or water for no template control reactions, to give a total reaction volume of 25 µl. Do not perform this step in the cabinet used for setting up reactions.
5. Transfer the plate or tubes to a suitable thermal cycling instrument, and run the following cycling conditions: 94°C for 2 min, followed by 40–45 three-step cycles of 94°C for 30 s, 58°C for 30 s and 72°C for 60 s, and a final extension step of 72°C for 5 min.
6. Cycling conditions should be optimized for the assay being used, the *Taq* polymerase used, the primer melting temperature and length of amplicon.
7. Clean up PCR products using the Qiagen QIAquick PCR Purification Kit following the manufacturer's instructions.
8. Transfer the DNA to a 0.6 ml tube and add 70 µl of Nexterion hybridization buffer.
9. Preheat the DNA at 95°C for 4 min, then cool to 55°C.
10. Condition the array tubes by adding 500 µl of water to the array tube and incubate in a thermal shaker at 55°C, 500 rpm for 5 min.
11. Remove the water and add 500 µl Nexterion hybridization buffer to the array tube and incubate in a thermal shaker at 55°C, 500 rpm for 5 min.

Protocol 7.2, Table 1. Composition of master mix.

Component[a]	Starting concentration	Volume per reaction (µl)	Final concentration
PCR buffer	5×	10	1×
$MgCl_2$	25 mM	7	3.5 mM
Biotin dCTP/dNTP mix[b]	4 mM total	4	0.32 mM total
Primer mix	Optimize for each multiplex	4	
GoTaq® polymerase	5 units/µl	0.25	0.025 units/µl
Water		To make up a final reaction volume of 49 µl	

[a]For detection of RNA targets, include reverse transcriptase, and an initial incubation at 48°C for 30 min before PCR cycling.
[b]Biotin dCTP/dNTP mix: 0.35 mM biotin dCTP, 0.65 mM dCTP, 1 mM dATP (deoxyadenosine triphosphate), 1 mM dGTP (deoxyguanosine triphosphate), 1 mM dTTP (deoxythymidine triphosphate).

12. Remove the Nexterion hybridization buffer from the array and add 100 µl of sample directly to the tube.

13. Incubate in a thermal shaker for 1 h at 55°C at 500 rpm.

14. Remove the buffer from the array and add 500 µl of wash buffer I. Incubate at 20°C, 500 rpm for 5 min.

15. Remove the buffer from the array and add 500 µl of wash buffer II. Incubate at 20°C, 500 rpm for 5 min.

16. Remove the buffer from the array and add 500 µl of wash buffer III. Incubate at 20°C, 500 rpm for 5 min.

17. Remove the buffer from the array and add 100 µl blocking solution. Incubate at 20°C for 15 min without shaking.

18. Remove the buffer from the array and add 100 µl conjugation solution. Incubate at 20°C for 15 min without shaking.

19. Remove the buffer from the array and add 500 µl PBS containing 0.1% Tween 20. Incubate at 20°C, 500 rpm for 5 min.

20. Repeat the wash a total of three times.

21. Remove buffer from array and add 100 µl SeramunGrün chip.

22. Clean external glass of the array tube.

23. Scan after 15–30 min.

24. For simple arrays, the Iconoclust software supplied with the ATR03 scanner will output a report of positive oligonucleotides and the intensity of the signal. For more complex arrays, the data can be exported as a comma limited file that can be used for more detailed analysis. For each feature on the array, a value for signal is given which is the intensity of that feature. The software also measures background and looks for unusual features that could be caused by dust or dirt, or non-specific signal. Such features are given a low confidence value (<0.7) or a '0' valid flag and should be removed from any analysis. An average of the signal for all the valid oligonucleotides for a particular virus can be calculated; this value should be significantly higher than the average of the negative control oligonucleotides. An ANOVA test with a Dunnett's post hoc test is suitable for this.

Notes

1. The typical size of a diagnostic array layout is 14×14 spots, but larger formats are available. Probes are generally spotted in duplicate.

2. Concentrations of primers should be optimized for each multiplex reaction.

3. Biotinylated DNA may be detected using HRP-linked streptavidin or HRP-linked antibiotin antibody; we have found the antibody to be more stable and to give better signals.

References

Barrett, B., Adams, I.P., Glover, R. and Boonham, N. (2009) The development of a Clondiag potato virus microarray for diagnostic use. *Aspects of Applied Biology* 94, 49–50.

Batchelor, M., Hopkins, K.L., Liebana, E., Slickers, P., Ehricht, R., Mafura, M., Aarestrup, F., Mevius, D., Clifton-Hadley, F.A., Woodward, M.J. *et al.* (2008) Development of a miniaturised microarray-based assay for the rapid identification of antimicrobial resistance genes in Gram-negative bacteria. *International Journal of Antimicrobial Agents* 31, 440–451.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. and Sayers, E.W. (2010) GenBank. *Nucleic Acids Research* 38, D46–D51.

Boonham, N., Walsh, K., Smith, P., Madagan, K., Graham, I. and Barker, I. (2003) Detection of potato viruses using microarray technology: towards a generic method for plant viral disease diagnosis. *Journal of Virological Methods* 108, 181–187.

Boonham, N., Tomlinson, J. and Mumford, R. (2007) Microarrays for rapid identification of plant viruses. *Annual Review of Phytopathology* 45, 307–328.

Bystricka, D., Lenz, O., Mraz, I., Dedic, P. and Sip, M. (2003) DNA microarray: parallel detection of potato viruses. *Acta Virologica* 47, 41–44.

Bystricka, D., Lenz, O., Mraz, I., Piherova, L., Kmoch, S. and Sip, M. (2005) Oligonucleotide-based microarray: a new improvement in microarray detection of plant viruses. *Journal of Virological Methods* 128, 176–182.

Cannon, G.A., Carr, M.J., Yandle, Z., Schaffer, K., Kidney, R., Hosny, G., Doyle, A., Ryan, J., Gunson, R., Collins, T. *et al.* (2010) A low density oligonucleotide microarray for the detection of viral and atypical bacterial respiratory pathogens. *Journal of Virological Methods* 163, 17–24.

Engel, E.A., Escobar, P.F., Rojas, L.A., Rivera, P.A., Fiore, N. and Valenzuela, P.D. (2010) A diagnostic oligonucleotide microarray for simultaneous detection of grapevine viruses. *Journal of Virological Methods* 163, 445–451.

Ghindilis, A.L., Smith, M.W., Schwarzkopf, K.R., Roth, K.M., Peyvan, K., Munro, S.B., Lodes, M.J., Stöver, A.G., Bernards, K., Dill, K. and McShea, A. (2007) CombiMatrix oligonucleotide arrays: genotyping and gene expression assays employing electrochemical detection. *Biosensors and Bioelectronics* 22, 1853–1860.

Julich, S., Riedel, M., Kielpinski, M., Urban, M., Kretschmer, R., Wagner, S., Fritzsche, W., Henkel, T., Möller, R. and Werres, S. (2011) Development of a lab-on-a-chip device for diagnosis of plant pathogens. *Biosensors and Bioelectronics* 26, 4070–4075.

King, A.M.Q., Adams, M.J., Carstens, E.B. and Lefkowitz, E.J. (eds) (2012) *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*. Academic Press (Elsevier), London, Waltham, Massachusetts and San Diego, California.

Ledford, H. (2008) The death of microarrays? *Nature* 455, 847.

Lee, G.P., Min, B.E., Kim, C.S., Choi, S.H., Harn, C.H., Kim, S.U. and Ryu, K H. (2003) Plant virus cDNA chip hybridization for detection and differentiation of four cucurbit-infecting tobamoviruses. *Journal of Virological Methods* 110, 19–24.

Monecke, S., Hochauf, K., Gottschlich, B. and Ehricht, R. (2006) A case of peritonitis caused by *Rhizopus microsporus*. *Mycoses* 49, 139–142.

Nicolaisen, M. (2011) An oligonucleotide-based microarray for detection of plant RNA viruses. *Journal of Virological Methods* 173, 137–143.

Pankhurst, L., Macfarlane-Smith, L., Buchanan, J., Anson, L., Davies, K., O'Connor, L., Ashwin, H., Pike, G., Dingle, K.E., Peto, T.E. *et al.* (2014) Can rapid integrated polymerase chain reaction-based diagnostics for gastrointestinal pathogens improve routine hospital infection control practice? A diagnostic study. *Health Technology Assessment* 18, 1–167.

Pasquini, G., Barba, M., Hadidi, A., Faggioli, F., Negri, R., Sobol, I., Tiberini, A., Caglayan, K., Mazyad, H., Anfoka, G. *et al*. (2008) Oligonucleotide microarray-based detection and genotyping of *Plum pox virus*. *Journal of Virological Methods* 147, 118–126.

Peplies, J., Glöckner, F.O. and Amann, R. (2003) Optimization strategies for DNA microarray-based detection of bacteria with 16S rRNA-targeting oligonucleotide probes. *Applied and Environmental Microbiology* 69, 1397–1407.

Schena, M., Shalon, D., Davis, R.W. and Brown, P.O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270, 467–470.

Tiberini, A. and Barba, M. (2012) Optimization and improvement of oligonucleotide microarray-based detection of tomato viruses and pospiviroids. *Journal of Virological Methods* 185, 43–51.

Tomlinson, J., Harrison, C., Boonham, N., Goodchild, S. and Weller, S. (2014) Influence of the length of target DNA overhang proximal to the array surface on discrimination of single-base mismatches on a 25-mer oligonucleotide array. *BMC Research Notes* 7, 251.

van Brunschot, S.L., Bergervoet, J.H., Pagendam, D.E., de Weerdt, M., Geering, A.D., Drenth, A. and van der Vlugt, R.A. (2014a) A bead-based suspension array for the multiplexed detection of begomoviruses and their whitefly vectors. *Journal of Virological Methods* 198, 86–94.

van Brunschot, S.L., Bergervoet, J.H., Pagendam, D.E., de Weerdt, M., Geering, A.D., Drenth, A. and van der Vlugt, R.A. (2014b) Development of a multiplexed bead-based suspension array for the detection and discrimination of pospiviroid plant pathogens. *PLoS ONE* 9(1), e84743.

Walsh, F., Cooke, N.M., Smith, S.G., Moran, G.P., Cooke, F.J., Ivens, A., Wain, J. and Rogers, T.R. (2010) Comparison of two DNA microarrays for detection of plasmid-mediated antimicrobial resistance and virulence factor genes in clinical isolates of Enterobacteriaceae and non-Enterobacteriaceae. *International Journal of Antimicrobial Agents* 35, 593–598.

Wang, D., Coscoy, L., Zylberberg, M., Avila, P.C., Boushey, H.A., Ganem, D. and Derisi, J.L. (2002) Microarray-based detection and genotyping of viral pathogens. *Proceedings of the National Academy of Sciences of the United States of America* 99, 15687–15692.

Wang, D., Urisman, A., Liu, Y.T., Springer, M., Ksiazek, T.G., Erdman, D.D., Mardis, E.R., Hickenbotham, M., Magrini, V., Eldred, J. *et al*. (2003) Viral discovery and sequence recovery using DNA microarrays 20. *PLoS Biology* 1, 257–260.

Watson, M., Dukes, J., Abu-Median, A.-B., King, D.P. and Britton, P. (2007) DetectiV: visualization, normalization and significance testing for pathogen-detection microarray data. *Genome Biology* 8, R190.

Wei, T., Pearson, M.N., Armstrong, K., Blohm, D. and Liu, J. (2012) Analysis of crucial factors resulting in microarray hybridization failure. *Molecular BioSystems* 8, 1325–1338.

Zhang, Y., Yin, J., Li, G., Li, M., Huang, X., Chen, H., Zhao, W. and Zhu, S. (2010) Oligonucleotide microarray with a minimal number of probes for the detection and identification of thirteen genera of plant viruses. *Journal of Virological Methods* 167, 53–60.

Zhang, Y., Yin, J., Jiang, D., Xin, Y., Ding, F., Deng, Z., Wang, G., Ma, X., Li, F., Li, G. *et al*. (2013) A universal oligonucleotide microarray with a minimal number of probes for the detection and identification of viroids at the genus level. *PLoS ONE* 8(5), e64474.

8 Next-generation sequencing

Rachel Glover* and Ian Adams

Fera, Sand Hutton, York, UK

8.1 Introduction

Next-generation sequencing (NGS) has revolutionized the discovery and sequencing of plant viruses. This chapter includes a description of how to generate sequences from a virus-infected plant and also presents protocols for the equally important post-sequencing quality control and bioinformatic analysis.

The genomics era of molecular biology started in 2004 with the publication of the human genome (Schmutz *et al.*, 2004), but this had taken 20 years and US$3 billion to complete on hundreds of individual DNA sequencers. In subsequent years, there was development that made DNA sequencing faster and cheaper. In 2005, the 454 GS 20 was the first of the 'next generation' machines to be released, and this was capable of producing 200,000 100 bp reads in a single run. This was 300 times the output of the equivalent sequencers used in the human genome project. The ability to produce large quantities of DNA sequence rapidly and relatively cheaply opened up the area of genomics to projects that would have previously been impossible but were now practical. Technical developments have continued, and the latest sequencer, the Illumina HiSeq X Ten, is capable of sequencing 18,000 human genomes in a year at a cost of US$1000 per genome.

The application of NGS to plant viruses began in 2009, when three different groups started to analyse virus-infected plants and, using different strategies to reduce the amounts of plant RNA in their samples, were able to sequence full-length viral genomes using NGS technologies: Adams *et al.* (2009) sequenced the total RNA on a Roche 454 FLX and used a subtractive hybridization approach to reduce the amount of non-viral plant RNA within the samples; Kreuze *et al.* (2009) used an Illumina GAII and

*Corresponding author: rachel.glover@fera.co.uk

sequenced small interfering RNA (siRNA; 23–24 nucleotide RNA molecules derived from viral RNA and produced by plants in response to viral infection); and Rwahnih *et al.* (2009) used a Roche 454 FLX to sequence double-stranded RNA (dsRNA) produced during viral replication. Since this time, NGS has become a recognized method for routine virus sequencing, and Barba *et al.* (2014) have written a useful review of its many applications.

The technology was recently used to identify the viruses responsible for a devastating disease that was destroying the maize harvest in Kenya in 2012 (Adams *et al.*, 2013). The viruses were identified as *Maize chlorotic mottle virus* and *Sugarcane mosaic virus*, which, in combination, are known to cause maize lethal necrosis. This achievement demonstrates one of the major benefits of NGS for plant virus identification. The strains of both viruses identified in Kenya were unusual, explaining why more targeted approaches such as ELISA had failed to detect them.

As stated earlier, a number of different approaches have been taken to reduce the amount of unnecessary sequencing of plant RNA and focus the sequencing effort on viral RNA. RNA has been the main focus for a majority of the studies, as most plant viruses have RNA genomes, and even DNA viruses have RNA replication intermediates. Of these methods, the most successful have been siRNA sequencing, reviewed by Kreuze (2014), and total RNA sequencing. This chapter will focus on the total RNA sequencing method. The choice of sequencing platform will depend on what is available, but Illumina (HiSeq and MiSeq) and the Roche 454 FLX platforms have been used successfully (Adams *et al.*, 2014; Barba *et al.*, 2014). While there are currently no publications that explicitly demonstrate the use of Ion Torrent (ThermoFisher Scientific) for use in plant virus discovery, there is no technical reason why this platform could not also be used.

8.2 Total RNA sequence generation

The basic total RNA sequencing method can be broken down into five stages: (i) sample extraction; (ii) library preparation; (iii) library normalization and pooling; (iv) sequencing; and (v) data analysis.

8.2.1 Sample extraction

Depending on the specific protocol being used, between 1 ng and 1 µg of DNA-free RNA is required to produce a sequencing library. RNA can be extracted using an RNeasy Kit with an on-column DNase digest (Qiagen). Other methods, such as those described in Chapter 2, which yield RT-PCR grade nucleic acid are also suitable. RNA can be quantified using spectrometry (NanoDrop, ThermoFisher Scientific) or fluorometry (Qubit, ThermoFisher Scientific).

It is now possible to remove plant ribosomal RNA, which can constitute up to 95% of the total extracted RNA. The RiboZero plant system (Illumina) uses plant ribosome-derived oligonucleotides attached to magnetic beads as capture probes, and can increase virus sequence yield tenfold (unpublished data).

8.2.2 Library preparation

Library preparation for all platforms (454, MiSeq/HiSeq, Ion Torrent) consists of fragmenting the RNA, converting the RNA to DNA using reverse transcriptase and then attaching DNA linkers with sequences required for the sequencing process. These linkers can also include sample-specific sequences called indexes (Illumina), multiplex identifiers (MIDs, Roche) or Ion barcodes (ThermoFisher Scientific).These allow multiple sample libraries to be combined and sequenced in a single run. Library preparation kits vary depending on the platform being used; for detailed protocols, refer to the manufacturer's instructions. For the Illumina MiSeq system, the ScriptSeq kit (Illumina) can be completed in 4 hours, and uses tagged random primers to create the first strand of DNA and a proprietary 'terminal tagging' RNA oligonucleotide to tag the 3′ end of the new single-stranded cDNA molecules. The kit then uses PCR with primers based on the tags to complete the second strand of DNA, add sample specific indexes and amplify the library.

8.2.3 Library normalization and pooling

Once constructed, the libraries need checking for quality and quantity, and if multiple libraries are being run, these need pooling in equimolar amounts. For library quality checking, using a Bioanalyser or TapeStation (Agilent) is effective, followed by quantification using fluorometry (Qubit, ThermoFisher Scientific) or real-time PCR (Kappa Biosystems). Figure 8.1 shows a typical TapeStation electropherogram for a ScriptSeq library. Libraries are pooled to give equimolar amounts of each library.

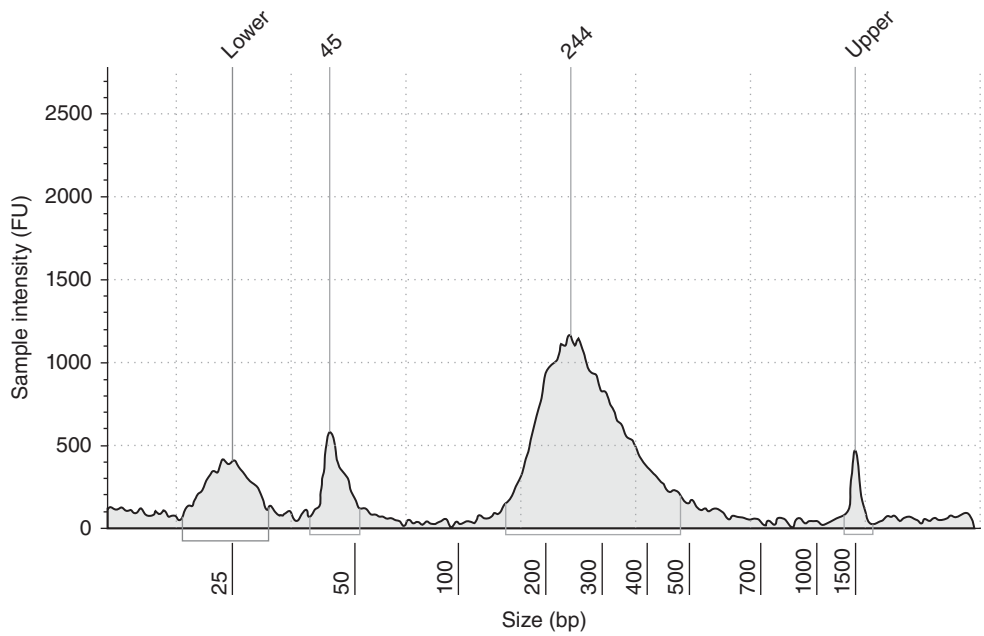


Fig. 8.1. Agilent TapeStation output of typical ScriptSeq library (pre-clean-up). The library has an average insert size of 244 bp. Sample intensity is given in fluorescence units (FU).

8.2.4 Sequencing

The specifics of sequencing a library will depend on the platform being used. For the 454 and Ion Torrent platforms, this consists of emulsion PCR, followed by bead recovery and plate loading. For the MiSeq platform, it consists of denaturing a diluted library, neutralizing it and then loading it on to the machine, where bridge amplification occurs automatically. Typical run times and yields are shown in Table 8.1.

Due to the high cost and complexity of equipment, it is frequently preferable to use a sequencing service provider.

8.2.5 Bioinformatics and analysis

NGS technologies have opened the door to many different analytical approaches and applications. Even the simplest of these analyses normally requires a level of computational proficiency and familiarity with UNIX-style servers, rather than being carried out on a typical Windows desktop computer. Protocol 8.2 assumes a basic familiarity with the Linux command line. For further information about this, there are some excellent resources available, including online tutorials, e.g. www.software-carpentry.org. The protocols supplied here are current for 2015. However, given that changes in NGS sequence analysis are extremely fast paced, there is no guarantee that these are future-proofed, although the principles of the analyses are likely to remain the same in any new tools and techniques that may be adopted in the future.

8.3 Quality control

Each NGS technology has different sequence data quality issues that are largely influenced by the chemistry used. The Roche 454 pyrosequencers have a tendency for homopolymer errors (Balzer *et al.*, 2011), Illumina sequencers for substitution errors (Nakamura *et al.*, 2011) and the Ion Torrent sequencer for insertion/deletion (indel) errors (Bragg *et al.*, 2013). A consideration of the types of errors introduced by the sequencing technology can be of benefit when interpreting the results of downstream analyses, and should be kept in mind.

8.3.1 Why perform quality control?

The raw data obtained from NGS sequencers is often screened for quality as part of the sequencing process. These procedures are typically described in the manufacturers' literature and are normally checks on the physical sequencing process, rather than

Table 8.1. Details of small-scale next-generation sequencing (NGS) platforms.

Platform	Supplier	Read length	Reads	Run time
454 GS-FLX+	Roche	750 bp	1 million	24 h
MiSeq v3	Illumina	2 × 300 bp	25 million	60 h
Ion Torrent 318	ThermoFisher Scientific	400 bp	5 million	7 h

on the true quality of the sequences being produced. Illumina, 454 and Ion Torrent sequencers all provide the sequences as FastQ files, which give both the DNA sequence and the quality scores for each base in the sequence. These files can be used for quality control and downstream analyses. Quality control for sequencing involves checking each sequence for poor-quality bases and drops in quality over the length of the sequence, and identifying sequencing adapters and control sequences. Trimming and removing poor-quality and contaminating sequences from the data set is then carried out to leave a 'clean' data set for downstream experiments.

Depending on the experiment being performed, quality control (QC) of sequence data can have a large impact on the results obtained. For example, without QC, the number of single nucleotide polymorphisms (SNPs) can be overestimated in genome sequences (as some will be sequencing errors), and this could be interpreted as a virus isolate having higher diversity or being more distantly related to other isolates. Though less likely, a particular concern is that of a poor sequencing run, where no post-sequencing QC has taken place before analysis. The end result could easily be misidentification of a pathogen due to sequencing errors or the incorrect conclusion that the isolate is a new species or subspecies.

8.3.2 Typical QC protocol

FASTQC for whole-data set evaluation

The most popular current software tool for assessing the quality of NGS sequences is FASTQC (Andrews, 2010). This tool is easy to use and can be run from a desktop computer or a server, depending on how large the data set is and it is desired to automate the QC of multiple sequence files. Further documentation is available on the FASTQC website.

FASTQC assesses 12 different aspects of sequence quality and gives each an overall 'pass/warning/fail' designation so the user can easily see which areas need further investigation (see Table 8.2). FASTQC assesses the quality of the sequences to give an overall picture of how good or poor the sequencing run is, but it should be noted that it does not trim sequences or remove poor sequences from the run data.

Removal of poor-quality and contaminating sequences

Once the quality of the overall data set has been assessed with FASTQC, there may be specific major problems with the sequences that will need to be addressed (e.g. large numbers of duplicated sequences or very short sequences). Addressing these sequence issues is outside the scope of this chapter, as many solutions are customized to the problem and data set to hand. However, all data sets without major issues should also be subjected to a basic screen for contaminating control and adapter sequences, and trimming to remove poor-quality sequences. The strategy below should be regarded as a starting point, depending on the type of sequence data and the experiments that are planned.

The initial task with any sequence data is to identify and remove any residual adaptor sequences. The manufacturer's software may state that this has already been carried out but, in practice, it is worth checking to remove any adaptors that may have slipped

Table 8.2. Sequence quality assessments carried out by FASTQC. Each aspect is summarized and the thresholds for triggering warnings and failure are described. (Adapted from the full FASTQC documentation available at www.bioinformatics.babraham.ac.uk/projects/fastqc/.)

Quality aspect	Summary	Warning	Failure
Basic statistics	Details the number of sequences, % GC (guanine, cytosine) content of all sequences, sequence lengths and the number of sequences flagged as poor quality	–	–
Per base sequence quality	Displays a graph showing the range of quality values along the length of all sequences in the data set	Median quality for any base <25	Median quality for any base <20
		Lower quartile quality for any base <10	Lower quartile quality for any base <5
Per tile sequence quality	Displays a quality plot heat map showing the quality along the length of all sequences against flow cell location (Illumina data only)	Mean Phred (quality) score more than 2 less than the mean for that base across all tiles	Mean Phred score more than 5 less than the mean for that base across all tiles
Per sequence quality scores	Displays a plot showing the quality score distribution for all sequences. Can show whether there is a subset of sequences of poorer quality	Most frequent mean quality <27	Most frequent mean quality <20
Per base sequence content	Displays a plot showing the proportion of each base at each position over all the sequences	Difference between proportions of ACTG (adenine, cytosine, thymine, guanine) is >10%	Difference between ACTG proportions is >20%

Per sequence GC content	Displays a plot showing the GC content at each base position over all the sequences on top of a modelled normal distribution	Sum of the deviations from the normal distribution represents >15% of all reads	Sum of the deviations from the normal distribution represents >30% of all reads
Per base N content	Displays a plot showing the percentage of 'N' (unknown) base calls at each position along the length of all sequences	Any position with >5% N content	Any position with >20% N content
Sequence length distribution	Displays a histogram of sequence lengths in the data set	If all the sequences are not the same length	If any sequences have a length of 0 bp
Sequence duplication levels	Displays a plot of the percentage of sequence duplication in the data set	>20% Non-unique sequences in the data set	>50% Non-unique sequences in the data set
Over-represented sequences	Lists all of the sequences that constitute >0.1% of the data set	Any sequence represents >0.1% of the data set	Any sequence representing >1% of the data set
Adapter content	Displays a plot of the cumulative proportion of the data set that has known adapter sequences present	Any adapter sequence represents >5% of the data set	Any adapter sequence representing >10% of the data set
Kmer (motifs of length k) content	Displays a plot of the proportions of calculated kmers across the sequence lengths	Any kmers showing bias with a *P* value <0.01	Any kmers showing bias with a *P* value <0.00001

through the net. This step will also need to be carried out if any kind of pre-amplification primers or custom adapters were used during library preparation. Following the removal of adaptor sequences, there is the further option of removing contaminating control sequences, in particular the PhiX control used in Illumina sequencing.

Removing and trimming poor-quality sequences from the data set is the critical QC step prior to downstream analyses. Most trimming software packages will use two approaches to trim poor-quality sequences. The first assesses the whole sequence and calculates the average Phred quality score for the entire read, discarding the read if the average is below a particular threshold. The second uses a 'sliding window' approach in which the quality is assessed in short stretches along the length of the sequence and, once it starts to drop, the sequence is trimmed. There are many different software tools that have been developed for adapter removal and quality trimming, and a selection of these are summarized in Table 8.3.

8.4 Metatranscriptomics for pathogen detection and viral whole genome sequencing

Metatranscriptomics, also referred to as RNA-Seq, is the large-scale sequencing of the total RNA complement of a sample and is typically carried out in order to determine differential gene expression (Mortazavi *et al.*, 2008; Wang *et al.*, 2009). In phytodiagnostics, it can also be a powerful tool for the detection and characterization of phytopathogens from host tissue (Adams *et al.*, 2009). Total RNA extraction and subsequent sequencing will not only detect bacterial and fungal pathogens present in the sample, but if any RNA viruses are present, their whole genomes will typically be sequenced, thus allowing detailed comparisons with other isolates to be carried out (Adams *et al.*, 2014).

8.4.1 Transcriptome assembly

The assembly of sequence reads from next-generation sequencers can be carried out in two ways: reference assembly (using an already-sequenced genome to align reads to) and *de novo* assembly (where no reference genome exists). For diagnostic purposes, the most sensible option is always *de novo* assembly of the transcriptome under investigation.

Table 8.3. List of software capable of removing adapter sequences and trimming poor-quality reads from sequencing data sets.

Software	Removes adapters?	Trims sequences?
Btrim (Kong, 2011)	Yes	Yes
Cutadapt (Martin, 2011)	Yes	No
Kraken (Davis *et al.*, 2013)	Yes	Yes
sickle (Joshi and Fass, 2011)	Yes	Yes
Trim Galore (Andrews, 2012)	Yes	Yes
Trimmomatic (Bolger *et al.*, 2014)	Yes	Yes

Plant genomes or transcriptomes used as a reference may be contaminated with microbial sequences labelled incorrectly as plant sequences in GenBank (the US National Institutes of Health genetic sequence database available from NCBI, the US National Center for Biotechnology Information). Using a reference data set with (unknown) contamination could potentially lead to a false negative result if only sequences that do not map to the host genome are analysed as potential phytopathogens. Our recommended approach to reduce false negatives and false positives is always to carry out *de novo* transcriptome assembly; this approach is described below.

Using Trinity to produce a de novo transcriptome assembly

Trinity (Grabherr *et al.*, 2011) is a transcriptome assembler that is run from the command line and can be used to assemble a transcriptome when no reference genome is available. Full details of all the options available and of the recommended hardware can be found at https://github.com/trinityrnaseq/trinityrnaseq/wiki. Given Illumina data and a server with adequate resources (>20 CPUs and >100 GB RAM), it is capable of assembling a transcriptome in a couple of hours. Servers with minimal resources will take days to weeks, depending on the amount of sequence data being assembled.

A basic transcriptome assembly command for Trinity is described in Box 8.1. Each transcript produced by Trinity has a unique name following the convention *cluster_gene_isoform*, so that c1_g3_i2 would be cluster 1, gene 3, isoform 2. The transcripts are grouped into clusters based on shared sequence content and these are loosely considered to be genes, so the true gene identifier for any transcript is the combination of the cluster and gene numbers (i.e. c1_g3 in the example given). For genes with multiple isoforms due to alternative splicing, for example, the isoform number gives an identifier to each variant.

Box 8.1. Trinity command for basic transcriptome assembly (more options are available).

```
trinity --seqType fq --left myLeftReads.fastq.gz --right myRightReads.fastq.gz --JM 400G --CPU 75 --bflyCPU 75 --output myassembly --full_cleanup
```

This command will produce an assembly using 400 GB RAM (--JM flag) and 75 CPUs (--CPU and -- bflyCPU flags), and the output transcriptome file in FastA format would be called myassembly.Trinity.fasta. The --full_cleanup flag removes intermediate files and folders, which can utilize large amounts of disk space and are normally not required for downstream analysis. Full details and options are available at https://github.com/trinityrnaseq/trinityrnaseq/wiki

Trinity also provides a very useful Perl script called TrinityStats.pl, which can be used to obtain basic information on the number of genes, number of isoforms, GC (guanine/cytosine) content and transcript length statistics. This can be useful in evaluating whether enough sequencing has been carried out to obtain acceptable coverage of the transcriptome being sequenced.

8.4.2 Sequence similarity searches

Once a transcriptome has been produced, the next step towards diagnosis of a pathogen is similarity searching of each transcript against a database of known sequences. This could be the entire GenBank database for true unknowns or a smaller database of pathogens for more routine screening.

Sequence similarity searching against the GenBank databases

The gold standard sequence similarity search for transcripts when attempting to identify phytopathogens is a blastx search against the GenBank non-redundant (nr) protein database. The large number of transcripts produced during this technique cannot be searched through the NCBI BLAST (Basic Local Alignment Search Tool) website (http://blast.ncbi.nlm.nih.gov/Blast.cgi), so we strongly recommend downloading the NCBI blast+ executables from ftp.ncbi.nlm.nih.gov/blast/executables/ and the entire 'nr' GenBank database formatted for blast searches from ftp.ncbi.nlm.nih.gov/blast/db/. These can then be installed on your local server (see the blast+ documentation for further details). The main blast+ programs are described in Table 8.4.

Table 8.4. Programs within the blast+ command line executables and used during sequence similarity searches for pathogens.

Program	Description
blastn	**Search a nucleotide database with a nucleotide query.** Megablast (high similarity searching) is the default, but true 'blastn' can be specified that searches for less similar sequences and is slower
blastp	**Search a protein database with a protein query.** A number of variants exist that extend the specificity of protein–protein searching, but blastp is the most widely used
blastx	**Search a protein database with a nucleotide query.** The first blast that should be used if the sample is infected with a suspected virus is blastx
tblastn	**Search a translated nucleotide database with a protein query.** Very rarely used in phytodiagnostics as open reading frames from GenBank nucleotide sequences are already present in the protein database searched with blastp and blastx
tblastx	**Search a translated nucleotide database with a translated nucleotide query.** Useful if searching for distantly related regions that are not open reading frames but could be present in GenBank as nucleotide sequences

The intended use of the blast searches downstream is to import them into software that will determine the taxonomic placement of each transcript and visualize the taxonomy. For this reason, there are a number of blast+ options that must be applied to any blast to aid the diagnostic specificity of the blast searches. A typical blastx search with diagnostics-specific options is shown in Box 8.2. This is particularly important when the suspected diagnosis is that of a plant virus. Other software programs, such as BLAT (Kent, 2002) and USEARCH (Edgar, 2010) are available, and these do a very good job at speeding up sequence similarity searches, although their search sensitivity is lower, which increases the likelihood of a false negative result.

Screening using a custom database of pathogen sequences

The creation of a custom blast database of pathogens of interest is easily achieved using the 'makeblastdb' program within the blast+ executables (Camacho *et al.*, 2009) that can be downloaded from NCBI (ftp.ncbi.nlm.nih.gov/blast/executables). However, searching for a specific list of pathogens will restrict the power of the approach in discovering new species and potentially novel isolates if the nucleotide or protein similarity is too far diverged from the species held in the database. In effect, this can result in false negatives if novel isolates are present. This should be taken into consideration when deciding on the trade-off between the speed of searching a

Box 8.2. Typical blastx search for downstream taxonomic analysis with MEGAN

```
blastx -db nr -query myassembly.trinity.fasta -out myassembly.trinity.blastx -num_descriptions 100 -num_alignments 100 -num_threads 75
```

This will produce a very large blast output file that can be used for downstream analysis. The 'num_descriptions' and 'num_alignments' flags ensure that an equal number of hits is described for descriptions and alignments, and at least 100 is recommended. The 'db' is the name of the database, in this case the GenBank 'nr' database, which should be installed locally. The 'num_threads' option is to tell blast+ how many CPUs to use in the search and is very important in speeding up the searching.

If you foresee that you might want to analyse the blast results with other software, we recommend including '-outfmt 5' to output the results as an XML file; however, this will greatly increase the size of the results file.

smaller screening database versus the greater sensitivity of the more comprehensive GenBank database.

8.4.3 Assigning taxonomy with MEGAN

Depending on the number of transcripts produced during assembly, the blast results file could be very large and challenging to process manually. If the results were outputted as XML it would be possible to assess the taxonomy by writing custom scripts in Python or Perl, but we would advise against reinventing the wheel and instead use MEGAN (Huson *et al.*, 2007). MEGAN (MEtaGenomeANalyzer) is a software tool that can be used on the desktop or from a server and can be used in metagenomic diagnostics to place taxonomically each transcript from a metagenomic sample and visualize the taxonomy across the whole sample. It can also be used to extract all of the sequences assigned to a particular species (or any other taxonomic level), which can be particularly useful for downstream analysis of any species of interest.

MEGAN assesses each transcript's blast hits to assign the transcript to a particular taxonomy and taxonomic level. The way it carries out this assessment is by ranking the blast hits by their score; the hits with a score that falls within a given percentage of the top hit score are selected (usually 10%). These blast hits are then used to calculate the lowest common ancestor (LCA) of that group of blast hits. As an illustration, using a recent example of novel viruses identified in carrot (Adams *et al.*, 2014), if all the blast hits selected for a transcript are *Carrot yellow leaf virus*, then the transcript will be given the species-level taxonomy of *Carrot yellow leaf virus*, as seen in Fig. 8.2. However, if the same transcript were to also contain a blast hit to a different species in the same genus (for example, *Citrus tristeza virus*), then the transcript would be assigned a genus-level taxonomy of *Closterovirus*. This type of assignment is particularly useful for discovering new species, as more than one species in a genus may appear with relatively equal blast scores to that of the transcript originating from the new species.

It is worth noting that MEGAN is entirely dependent on the NCBI taxonomy database and assumes that the taxonomy assigned to sequences in the GenBank database is

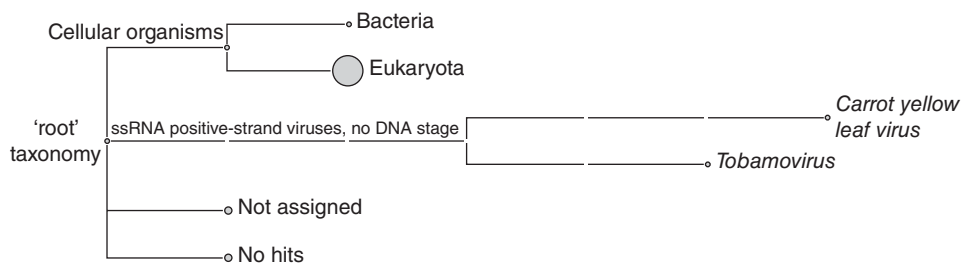


Fig. 8.2. Typical MEGAN taxonomy visualization of a metagenomic sample showing the presence of *Carrot yellow leaf virus* and a tobamovirus. (Modified from Adams *et al.*, 2014.)

correct. In our experience, there are many novel plant virus sequences from historical EST (expressed sequence tag) sequencing that have been assigned plant taxonomies within GenBank. When the lowest common ancestor algorithm is applied to blast results that contain these 'plant' sequences, an LCA between plant and virus sequences is calculated that results in these transcripts being assigned to the root of the taxonomy. If the aim of an experiment is to search for novel plant viral isolates and there are no immediate candidates in the virus taxonomy of MEGAN, they could be affected by one of these rogue plant sequences and will be found in the 'root' section of the visualization.

Protocols

Contents

Protocol 8.1

Plant viral sequence generation using ScriptSeq libraries and an Illumina MiSeq

Materials

- Illumina MiSeq (or access through a service provider)
- Agilent TapeStation or Bioanalyser
- Thermal cycler
- Pipettors and filter tips (separate sets for pre- and post-PCR stages)
- Magnetic separation stand
- Microfuge
- PicoGreen/Qubit double-stranded DNA fluorometer and reagents (ThermoFisher Scientific)
- AMPure XP magnetic beads (Agencourt)
- RNAclean XP magnetic clean-up beads (Agencourt)
- ScriptSeq Complete (plant leaf or root) library preparation kit (Illumina)
- ScriptSeq index kit (Illumina)
- Failsafe DNA polymerase (Illumina)
- PCR plates, Eppendorf tubes
- Sodium hydroxide (10 N)
- PhiX control library (Illumina)
- 1 µg DNA-free RNA from infected plant
- MiSeq V3 600 cycle run kit, including hybridization buffer (HT1) (Illumina)

Methods

1. Follow the instructions in the ScriptSeq Complete manual to:
   (i) ribosome deplete the RNA with RiboZero reagents (Illumina)
   (ii) clean up RNA with RNA clean;
   (iii) make first strand cDNA;
   (iv) terminally tag with linkers;
   (v) clean with Ampure XP beads;
   (vi) PCR amplify and add indexes; and
   (vii) clean up with Ampure XP.
2. Assess quality of the library with TapeStation/Bioanalyser and quantify with Qubit/PicoGreen.
3. Follow the instruction in the MiSeq manual to:
   (i) dilute library to 4 nM;
   (ii) denature with sodium hydroxide;
   (iii) neutralize with HT1 buffer;
   (iv) dilute to an appropriate concentration (8–15 pM) and mix with 1% denatured PhiX library (20 pM);
   (v) load into MiSeq reagent cartridge;
   (vi) load MiSeq with flow cell and reagent cartridge; and
   (vii) run MiSeq.

Protocol 8.2

Plant viral sequence analysis

Materials

- Multicore Linux server with large amounts of RAM. (We use 60 core 500 GB RAM server running Fedora. It is also possible to use Amazon cloud, although installation of software is time-consuming.)
- Software installed (see main text for links):
  - FastQC
  - sickle
  - Trinity
  - blast+
  - NCBI BLAST nr/nt databases
  - MEGAN

Methods

1. Quality check data using FASTQC.
2. Trim poor-quality data using sickle.
3. Assemble using Trinity.
4. blastx search against NCBI BLAST nr database.
5. Visualize, identify and extract viruses using MEGAN.

References

Adams, I.P., Glover, R.H., Monger, W.A., Mumford, R., Jackeviciene, E., Navalinskiene, M., Samuitiene, M. and Boonham, N. (2009) Next-generation sequencing and metagenomic analysis: a universal diagnostic tool in plant virology. *Molecular Plant Pathology* 10, 537–545.

Adams, I.P., Miano, D.W., Kinyua, Z.M., Wangai, A., Kimani, E., Phiri, N., Reeder, R., Harju, V., Glover, R., Hany, U. *et al.* (2013) Use of next-generation sequencing for the identification and characterization of *Maize chlorotic mottle virus* and *Sugarcane mosaic virus* causing maize lethal necrosis in Kenya. *Plant Pathology* 62, 741–749.

Adams, I.P., Skelton, A., Macarthur, R., Hodges, T., Hinds, H., Flint, L., Deb Nath, P., Boonham, N. and Fox, A. (2014) *Carrot yellow leaf virus* is associated with carrot internal necrosis. *PLoS ONE* 9(11), e109125.

Andrews, S. (2010) FastQC. A quality control tool for high throughput sequence data. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc (accessed 4 November 2015).

Andrews, S. (2012) Trim Galore!: A wrapper tool around Cutadapt and FastQC to consistently apply quality and adaptor trimming to fastQ files, with some extra functionality for MspI-digested RRBS-type (Reduced Representation Bisufite-Seq) libraries. Available at: http://www.bioinformatics.babraham.ac.uk/projects/trim_galore (accessed 4 November 2015).

Balzer, S., Malde, K. and Jonassen, I. (2011) Systematic exploration of error sources in pyrosequencing flowgram data. *Bioinformatics* 27, i304–i309.

Barba, M., Czosnek, H. and Hadidi, A. (2014) Historical perspective, development and applications of next-generation sequencing in plant virology. *Viruses* 6, 106–136.

Bolger, A.M., Lohse, M. and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120.

Bragg, L.M., Stone, G., Butler, M.K., Hugenholtz, P. and Tyson, G.W. (2013) Shining a light on dark sequencing: characterising errors in Ion Torrent PGM data. *PLoS Computational Biology* 9(4), e1003031.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T.L. (2009) Blast+: architecture and applications. *BMC Bioinformatics* 10, 421.

Davis, M.P., van Dongen, S., Abreu-Goodger, C., Bartonicek, N. and Enright, A.J. (2013) Kraken: a set of tools for quality control and analysis of high-throughput sequence data. *Methods* 63, 41–49.

Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q. *et al.* (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29, 644–652.

Huson, D.H., Auch, A.F., Qi, J. and Schuster, S.C. (2007) MEGAN analysis of metagenomic data. *Genome Research* 17, 377–386.

Joshi, N.A. and Fass, J.N. (2011) sickle – A windowed adaptive trimming tool for FastQ files using quality. Available at: https://github.com/najoshi/sickle (accessed 4 November 2015).

Kent, W.J. (2002) BLAT – the BLAST-like alignment tool. *Genome Research* 12, 656–664.

Kong, Y. (2011) Btrim: a fast, lightweight adapter and quality trimming program for next-generation sequencing technologies. *Genomics* 98, 152–153.

Kreuze, J. (2014) siRNA deep sequencing and assembly: piecing together viral infections. In: Gullino, M.L. and Bonants, P.J.M. (eds) *Detection and Diagnostics of Plant Pathogens*. Springer, Dordrecht, the Netherlands, pp. 21–38.

Kreuze, J.F., Perez, A., Untiveros, M., Quispe, D., Fuentes, S., Barker, I. and Simon, R. (2009) Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: a generic method for diagnosis, discovery and sequencing of viruses. *Virology* 388, 1–7.

Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMB.net.journal* 17, 10–12.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. and Wold, B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* 5, 621–628.

Nakamura, K., Oshima, T., Morimoto, T., Ikeda, S., Yoshikawa, H., Shiwa, Y., Ishikawa, S., Linak, M.C., Hirai, A., Takahashi, H. *et al*. (2011) Sequence-specific error profile of Illumina sequencers. *Nucleic Acids Research* 39, e90.

Rwahnih, M., Daubert, S., Golino, D. and Rowhani, A. (2009) Deep sequencing analysis of RNAs from a grapevine showing Syrah decline symptoms reveals a multiple virus infection that includes a novel virus. *Virology* 387, 395–401.

Schmutz, J., Wheeler, J., Grimwood, J., Dickson, M., Yang, J., Caoile, C., Bajorek, E., Black, S., Chan, Y.M., Denys, M. *et al*. (2004) Quality assessment of the human genome sequence. *Nature* 429, 365–368.

Wang, Z., Gerstein, M. and Synder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* 10, 57–63.

9 On-site testing for plant pathogens

Jenny Tomlinson*

Fera, Sand Hutton, York, UK

9.1 Introduction

Testing for plant pests and pathogens may be done for a number of different reasons, and the criteria by which alternative methods are judged and selected are influenced by the context. In most cases, primary concerns are the accuracy of identification and the limit of detection: in other words, the specificity and sensitivity of the test. The performance characteristics of a test can be considered in terms of what levels of false positive and false negative results might be acceptable, or the required limit of detection in light of the samples to be tested. Where the number of samples to be tested is very high, additional key factors relate to throughput and the cost per sample. For applications such as surveying or routine monitoring, the preferred methods may be those that can be scaled up for large numbers of samples and wholly or partially automated, such as ELISA or real-time PCR.

In some situations, a significant consideration in assessing the value of a testing method is the time it takes to achieve a result. For example, where the objective of testing is to determine whether imported commodities are free from quarantine pests and pathogens, the time taken to obtain a result is critical. For the most part, testing for pests and pathogens is carried out in dedicated laboratory facilities, and there is an inherent delay while samples are in transit to the laboratory. However, this delay can be eliminated by carrying out testing at the point of sampling.

Methods suitable for use under non-laboratory conditions differ in a number of ways from the methods developed for high-throughput testing in potentially sophisticated laboratory facilities. Ideal methods for non-laboratory testing are rapid, simple to perform (i.e. they involve as few manipulations as possible) and require either very

*E-mail: jenny.tomlinson@fera.co.uk

simple equipment or none at all. Any instruments required should be robust and have low power requirements (or run off batteries), and may, therefore, be portable as well. A lack of laboratory support facilities may limit options for the decontamination of equipment and waste disposal, so methods may be favoured that use non-toxic, disposable consumables but, conversely, do not generate excessive amounts of waste.

Methods developed for this kind of on-site testing application are also likely to be more suitable than conventional laboratory procedures for use in low-resource settings, e.g. basic 'field laboratories'. In summary, the aim of developing non-laboratory detection methods is to allow testing to be performed under conditions where time constraints or a lack of resources, or a combination of these factors, would otherwise either preclude testing entirely, or impose the use of tests with suboptimal performance characteristics.

9.2 On-site testing using antibody-based methods

One method for the detection of plant pathogens that is very well suited to on-site testing, and which has been deployed with much success for a range of targets, is the use of immunoassays in a lateral flow device (LFD) format (Danks and Barker, 2000). These tests use the same technology as a double antibody sandwich (DAS) ELISA. The sample to be tested is partially homogenized in phosphate buffer (typically by shaking in a bottle containing ball bearings) before being applied to a nitrocellulose membrane impregnated with the reagents for a target-specific sandwich immunoassay, with a simple visual read-out in which development of a coloured line represents a positive test (see Fig. 9.1). Tests in an LFD format are in many ways an ideal solution for on-site testing, as they are extremely simple to perform, rapid (taking <5 min) and require relatively low-cost, disposable consumables. LFDs are valuable tools for the detection of a range of plant pathogens (Thornton *et al.*, 2004; Lane *et al.*, 2007; Braun-Kiewnick *et al.*, 2011). There are two main limitations to the use of LFDs. The first is that for some targets, the sensitivity and/or specificity is insufficient for testing in certain scenarios; these factors are both

Fig. 9.1. Lateral flow device (LFD) for pathogen detection.

determined by the properties of the available antibodies for a given target. Second, while highly sensitive and specific antibodies have been developed for some plant pathogens, the development of antibodies for a given target can be time-consuming (taking a minimum of several months) and technically challenging for some organisms. For example, antibodies developed for the detection of fungal pathogens may be life-stage specific, may lack species-level specificity or may be insufficiently sensitive to detect low-titre or asymptomatic infections. For these reasons, while LFDs in many respects represent an ideal example of an on-site testing method, for some targets, the development of simplified nucleic acid-based methods is the preferred approach to on-site testing.

As described in previous chapters, detection methods based on amplification of nucleic acid offer high degrees of sensitivity and specificity. For use in non-laboratory conditions, the challenge is to simplify these methods to increase the speed of testing, reduce the logistic burden and, crucially, to maximize the robustness of the methods for use by different end-users in a range of scenarios.

9.3 On-site nucleic acid extraction

As described in Chapter 2, the first step in nucleic acid-based tests is the extraction of nucleic acid from the sample matrix. In the laboratory, these methods can be relatively time-consuming and often involve the use of relatively complex laboratory equipment. Even the simplest methods typically require centrifugation and/or incubation of samples at specific temperatures. The two main considerations for nucleic acid extraction under non-laboratory conditions are the need to disrupt the sample matrix sufficiently to release amplifiable DNA, and the potential for inhibitory substances to be released from the sample. If significant amounts of inhibitors are released, these may need to be removed before amplification, or their effects on downstream processes (such as PCR) need to be mitigated by some other means.

Conventional methods for the extraction of nucleic acid from plant material are generally based on the principle that the best yield will be obtained by homogenizing the sample as completely as possible, after which clean-up steps are carried out to remove potentially inhibitory substances co-extracted with the DNA. These steps often require the use of chemicals or equipment that are largely restricted to use in the laboratory (e.g. precipitation and centrifugation or processes using silica beads or membranes for separation). An alternative approach to extraction under non-laboratory conditions is to disrupt the sample material only partially, so that fewer inhibitory substances are released and additional clean-up steps are not required. An example of this approach uses alkaline polyethylene glycol (PEG) buffer (Chomczynski and Rymaszewski, 2006) to produce a crude extract which can be tested directly without additional steps. The use of crude extraction methods is facilitated by the use of amplification chemistries or reagents that exhibit enhanced tolerance of potentially inhibitory compounds.

An alternative approach to simplified extraction involves stabilizing the nucleic acid on a solid matrix. Whatman FTA technology is the best known example of this approach (Ndunguru *et al.*, 2005). Crude samples are applied to cards, which are then chemically treated to lyse cells and capture nucleic acid within the fibres of the matrix. This method

has the advantage that the nucleic acid is stabilized on the card, thus allowing samples to be transported and stored at room temperature. Sections can then be punched out of the cards and washed before being added directly to amplification reactions. A similar workflow can be used for the stabilization and subsequent testing of nucleic acid using flow-through nitrocellulose membranes (Tomlinson *et al.*, 2010).

9.4 Real-time PCR in the field

An approach to the use of nucleic acid-based detection methods in the field is to modify established laboratory methods for non-laboratory conditions. Real-time PCR has advantages of speed, sensitivity and specificity over other methods, including conventional PCR, and is increasingly used for routine, high-throughput testing. Real-time PCR is mostly restricted to laboratory use by the complexity of the equipment required for concurrent thermal cycling and fluorescence monitoring, although portable and ruggedized real-time PCR instruments have been developed and used for the detection of plant pathogens under non-laboratory conditions. The Cepheid SmartCycler (Fig. 9.2) allows 16 samples to be tested by real-time PCR, with run times of around an hour. This instrument can be used in conjunction with simplified extraction methods for the detection of plant pathogens outside central laboratories (Schaad *et al.*, 2002; Tomlinson *et al.*, 2005; Mumford *et al.*, 2006). However, as noted above, real-time PCR generally requires some clean-up of nucleic acid for detection to be reliable. Instruments, including the Cepheid GeneXpert and the BioFire Diagnostics FilmArray, have been developed to perform integrated sample preparation and real-time PCR for clinical samples, but instrumentation and per-test costs currently preclude the use of these systems for routine plant health applications. Conventional nucleic acid extraction methods using magnetic silica beads can be modified, for example, by using a

Fig. 9.2. The Cepheid SmartCycler for real-time PCR.

handheld magnetic device such as the PickPen (Bio-Nobile) to enable extraction to be carried out without laboratory equipment (Tomlinson *et al.*, 2005; Hughes *et al.*, 2006). This approach can produce extracts that are suitable for PCR-based testing, but this is still a multistep process, making it suitable for use in some low-resource settings but not all on-site applications.

A potential limitation of PCR-based testing, particularly in comparison with tests such as those done in an LFD format, is the time taken to obtain a result. Fast-cycling PCR is technically achievable, with run times which are comparable to those of isothermal methods, including loop-mediated isothermal amplification (LAMP), but rapid cycling conditions have been observed to have an adverse effect on PCR sensitivity (Hilscher *et al.*, 2005), and may exacerbate the demand for very clean DNA.

9.5 Isothermal amplification in the field

As described in Chapter 4, isothermal amplification methods including LAMP present alternatives to PCR that have a number of potential advantages. In the context of on-site testing, the most important of these are the rapid amplification times (<30 min) and reduced susceptibility to inhibition. LAMP is sufficiently tolerant of inhibitors originating from the host matrix to be used in conjunction with very simple crude extraction procedures, for example using alkaline PEG buffer (Protocol 9.2).

LAMP can be used in conjunction with a range of end-point detection methods, some of which are particularly suitable for use in the field. Due to the large amount of product generated in a LAMP reaction, the difference between a negative and a positive reaction is generally unambiguous, so colour change reactions can be used to give a simple readout of results, which is ideal for non-specialist users. LFDs for the detection of labelled products (Chapter 4) also have the advantage of providing a readout which is unambiguous and easy to interpret. The large amount of DNA after amplification using LAMP is also a potentially significant disadvantage due to the risk of carry-over contamination, which is compounded in the case of non-expert operators. The risk of contamination can be reduced by minimizing the number of manipulations required to perform the test, and by the separation of pre- and post-amplification steps. Real-time LAMP using fluorescence detection can also be used under non-laboratory situations; the closed-tube nature of real-time detection greatly reduces the risk of carry-over contamination, and the melt/anneal analysis that can be performed increases the robustness of results. This is particularly valuable when testing crude extracts, when the nature and condition of the sample could increase the likelihood of both false positive and false negative results. Instruments, including the OptiGene Genie II (Fig. 9.3), have been developed for real-time isothermal amplification. The Genie II is particularly suitable for non-laboratory use owing to its small size and portability, and its ability to run on battery power for several hours.

9.6 Deployment of tests in the field

The end-users of a method intended for on-site testing are likely to differ significantly from the users of laboratory tests in terms of their particular skills and degree of specialism, and this difference should be taken into account to allow on-site tests to be

Fig. 9.3. The OptiGene Genie II instrument for real-time loop-mediated isothermal amplification (LAMP).

deployed effectively in the field. To facilitate testing under a range of conditions by a variety of end-users, and with prompt availability of results, tests should ideally involve the smallest possible number of hands-on steps, in particular, minimizing the number of specialist procedures such as pipetting small volumes of liquids, even if these are routine in the laboratory. By selecting methods that conform to these requirements, uptake by end-users is likely to be increased, as testing can be carried out more easily in conjunction with other activities, such as the visual inspection of symptoms. As well as simplicity in carrying out the test, it is also preferable that the results obtained should be simple to interpret without ambiguity. In pursuit of this objective, and with the aspiration of replicating laboratory-quality results in the field, it will generally be necessary to incorporate suitable controls into the testing procedure. Where possible, negative and positive control reactions should be run to rule out false positive and false negative results, respectively, and parallel testing with an internal control assay can help to identify samples for which testing has failed. Validation to compare the results of the on-site method with laboratory-based 'gold standard' methods can give the end-user confidence in the degree to which the result of the on-site test has replicated the result that would have been obtained in the laboratory.

On-site testing may be particularly attractive as a screening tool to select samples to be sent to the laboratory for further testing, and relatively cheap and simple methods are required for this to be worthwhile. Depending on the context, false negative or positive results may be particularly undesirable. If screening is carried out with the aspiration of detecting all positive samples, the screening method needs to be highly sensitive to avoid false negatives; DNA-based tests are particularly desirable for this reason. A highly sensitive screening method could be used in conjunction with the laboratory confirmation of positives if the field method is insufficiently specific; however, species-level specificity is routinely achievable using methods such as LAMP, with the potential to target other taxonomic levels if required. The robustness of testing may be improved

in some cases by testing using multiple methods, which may be carried out sequentially and include both field and laboratory methods. Commonly, the decision of whether or not to use each method will depend on the results of the previous test; for example, visual inspection of samples with symptoms may lead to on-site testing if the symptoms conform to the expectations for the disease (or otherwise raise suspicion sufficiently to justify testing), with the sample being submitted to the laboratory for confirmation on the basis of the result. Assessment of the diagnostic sensitivity and specificity of individual methods allows the calculation of predictive values (Altman and Bland, 1994; López *et al*., 2009), which apply in particular scenarios, and allow the user to determine whether it is necessary to confirm either positive or negative results by additional testing. In a particular scenario, it is, therefore, necessary to consider the performance characteristics of each test (sensitivity, specificity), the relevant logistical factors, including cost and time to result, and the ultimate objective of testing (e.g. eradicating a disease, preventing the entry of a disease to an area, assessing the distribution of a pathogen in an area, or testing in order to recommend disease control measures) in order to determine the optimal testing regime.

Protocols

Contents

Protocol 9.1

Real-time PCR using the Cepheid SmartCycler

Materials

- Reagents for real-time PCR (see Chapter 3): *Taq* DNA polymerase, PCR buffer (usually supplied with *Taq* polymerase), $MgCl_2$, dNTPs (deoxynucleotide triphosphates), TaqMan primers and probe, molecular biology grade water
- SmartCycler reaction tubes (Cepheid)
- Pipettors – separate sets for setting up the reactions and adding DNA to the reactions
- Filter tips for pipettors
- DNA extracted using a method suitable for on-site use[1]
- SmartCycler instrument (Cepheid) and laptop computer
- Microcentrifuge with rotor for spinning SmartCycler reaction tubes

Method

1. Real-time PCR reactions can be set up in advance and kept on ice for several hours before use. The real-time PCR master mix should be set up as described in Chapter 3, following procedures to avoid contamination.
2. Pipette 24 µl aliquots of master mix into SmartCycler reaction tubes, partially close the lids and store on ice until ready to use.
3. Add 1 µl DNA extract to each tube and close the lids firmly. Include at least one negative control reaction in each run in which DNA is replaced with molecular biology grade water. Spin the tubes in the microcentrifuge.

4. Place the tubes into the SmartCycler. Up to 16 reactions can be run at one time.
5. Run the instrument using the required cycling conditions (see Chapter 3). Typical cycling conditions for TaqMan real-time PCR (95°C for 10 min followed by 40 cycles of 95°C for 15 s and 60°C for 1 min) will be completed in approximately 70 min.
6. Interpret the results in terms of Ct values (the cycle at which the fluorescent signal crosses the threshold value), as described in Chapter 3.

Note

1. The chosen method will depend on the particular conditions under which testing is to be carried out. Extraction kits that can be used with little or no equipment include the ChargeSwitch gDNA Plant Kit (ThermoFisher Scientific) and the QuickPick Plant DNA kit (Bio-Nobile). Both of these kits are based on the manipulation of magnetic beads, which can be achieved using a handheld device such as the PickPen or QuickPick magnetic tool (Bio-Nobile).

Protocol 9.2

PEG extraction and real-time LAMP using the OptiGene Genie II

Materials

- PEG (polyethylene glycol) buffer[1]
- 5 ml screw cap tubes
- Stainless steel ball bearings (7/16″ diameter)[2]
- Molecular biology grade water
- Pipettors – separate sets for setting up the reactions and adding DNA to the reactions[3]
- Filter tips for pipettors
- Microcentrifuge tubes
- Reagents for real-time LAMP (loop-mediated isothermal amplification; see Protocol 4.2)
- Genie reaction tubes (OptiGene)
- Genie II instrument (OptiGene)

Method

1. Samples of plant material can be prepared by shaking for 1–2 min in a screw-cap tube containing 1 ml PEG buffer and a ball bearing.
2. Dilute the PEG extract by a factor of 1 in 10 (for example, transfer 10 µl PEG extract into a tube containing 90 µl molecular biology grade water and shake or invert to mix).
3. Real-time LAMP reactions can be set up in advance and kept on ice for several hours before use. The real-time LAMP master mix should be set up as described in Protocol 4.2, following procedures to avoid contamination.
4. Pipette 24 µl aliquots of master mix into Genie reaction tubes (in 8 well strips), partially close the lids and store on ice until ready to use.
5. Add 1 µl diluted PEG extract to each tube and close the lids firmly. Include at least one negative control reaction in each run in which DNA is replaced with molecular biology grade water.

6. Place the tubes into the Genie II. Up to 16 reactions can be run at one time.
7. Run the instrument using the required conditions (see Protocol 4.2). Interpret the results in terms of Tp values (the time taken to reach a maximum rate of fluorescence increase) and anneal peaks (see Chapter 4).

Notes

1. PEG buffer: 60% PEG 200 (Sigma) plus 20 mM KOH (or NaOH), pH 13.3–13.5.
2. This method has been found to be effective for a range of plant tissues, including leaf material and small pieces of wood. Other methods can also be used; note that complete homogenization is not required.
3. Dilution can be carried out using a 10 µl disposable plastic inoculating loop instead of a pipettor.

References

Altman, D.G. and Bland, J.M. (1994) Diagnostic tests 2: Predictive values. *The BMJ* 309, 102.

Braun-Kiewnick, A., Altenbach, D., Oberhänsli, T., Bitterlin, W. and Duffy, B. (2011) A rapid lateral-flow immunoassay for phytosanitary detection of *Erwinia amylovora* and on-site fire blight diagnosis. *Journal of Microbiological Methods* 87, 1–9.

Chomczynski, P. and Rymaszewski, M. (2006) Alkaline polyethylene glycol-based method for direct PCR from bacteria, eukaryotic tissue samples, and whole blood. *BioTechniques* 40, 454–458.

Danks, C. and Barker, I. (2000) On-site detection of plant pathogens using lateral-flow devices. *EPPO Bulletin* 30, 421–426.

Hilscher, C., Vahrson, W. and Dittmer, D.P. (2005) Faster quantitative real-time PCR protocols may lose sensitivity and show increased variability. *Nucleic Acids Research* 33, e182.

Hughes, K.J.D., Giltrap, P.M., Barton, V.C., Hobden, E., Tomlinson, J.A. and Barber, P. (2006) On-site real-time PCR detection of *Phytophthora ramorum* causing dieback of *Parrotia persica* in the UK. *Plant Pathology* 55, 813.

Lane, C.R., Hobden, E., Walker, L., Barton, V.C., Inman, A.J., Hughes, K.J.D., Swan, H., Colyer, A. and Barker, I. (2007) Evaluation of a rapid diagnostic field test kit for identification of *Phytophthora* species including *P. ramorum* and *P. kernoviae* at the point of inspection. *Plant Pathology* 56, 828–835.

López, M.M., Llop, P., Olmos, A., Marco-Noales, E., Cambra, M. and Bertolini, E. (2009) Are molecular tools solving the challenges posed by detection of plant pathogenic bacteria and viruses? *Current Issues in Molecular Biology* 11, 13–46.

Mumford, R., Boonham, N., Tomlinson, J. and Barker, I. (2006) Advances in molecular phytodiagnostics – new solutions for old problems. *European Journal of Plant Pathology* 116, 1–19.

Ndunguru, J., Taylor, N.J., Yadav, J., Aly, H., Legg, J.P., Aveling, T., Thompson, G. and Fauquet, C.M. (2005) Application of FTA technology for sampling, recovery and molecular characterization of viral pathogens and virus-derived transgenes from plant tissues. *Virology Journal* 2, 45.

Schaad, N.W., Opgenorth, D. and Gaush, P. (2002) Real-time polymerase chain reaction for one-hour on-site diagnosis of Pierce's disease of grape in early season asymptomatic vines. *Phytopathology* 92, 721–728.

Thornton, C.R., Groenhof, A.C., Forrest, R. and Lamotte, R. (2004) A one-step, immunochromatographic lateral flow device specific to *Rhizoctonia solani* and certain related species, and its use to detect and quantify *R. solani* in soil. *Phytopathology* 94, 280–288.

Tomlinson, J.A., Boonham, N., Hughes, K.J.D., Griffin, R.L. and Barker, I. (2005) On-site DNA extraction and real-time PCR for detection of *Phytophthora ramorum* in the field. *Applied and Environmental Microbiology* 71, 6702–6710.

Tomlinson, J.A., Dickinson, M., Hobden, E., Robinson, S., Giltrap, P.M. and Boonham, N. (2010) A five-minute DNA extraction method for expedited detection of *Phytophthora ramorum* following prescreening using *Phytophthora* spp. lateral flow devices. *Journal of Microbiological Methods* 81, 116–120.

10 Quality assurance for molecular testing in plant health

Aude Chabirand,[1]* Geraldine Anthoine,[2] Adrian Fox[3] and Lynn Laurenson[3]

[1]*Plant Health Laboratory, Anses (French Agency for Food, Environmental and Occupational Health and Safety), Saint Pierre 97410, Réunion, France;* [2]*Plant Health Laboratory, Anses (French Agency for Food, Environmental and Occupational Health and Safety), Angers, France;* [3]*Fera, Sand Hutton, York, UK*

10.1 Introduction

An accurate diagnosis in the field of plant health is a prerequisite to the effective control of plant pathogens. Quality assurance for molecular testing consists of activities that ensure the quality and confidence of a diagnosis performed in a molecular laboratory. It is based both on management and technical requirements.

Molecular biology is playing a growing part in plant health laboratories, so the development of quality management systems for molecular testing has become a concern for many laboratories. The requirements for accreditation to the international standard are given in the International Organization for Standardization/International Electrochemical Commission standard ISO/IEC 17025:2005 (ISO/IEC, 2005). Accreditation for molecular testing in plant health is a formal recognition that a testing laboratory is competent to carry out specific molecular tests to detect plant pathogens.

*Corresponding author: aude.chabirand@anses.fr

The implementation of a quality approach in a molecular laboratory is not really any different from the implementation of a quality approach in a laboratory performing analysis based on other biological disciplines, but one of the main obstacles to this implementation is lack of understanding of this approach. Such an attitude is all the more regrettable given that the laboratory has everything to gain from providing test results in which both its clients and the laboratory itself can be confident.

This chapter reviews the key elements for the implementation of quality assurance for molecular testing in plant health, including test selection and validation, staff training, facilities, supplies, estimation of measurement uncertainty, documental traceability, quality control and continuous improvement.

10.2 Fit for purpose testing

Too often, the tendency is to describe an analytical method as a fixed and unchanging procedure. However, like any production process, an analytical method is born, evolves and dies. The life of a method comprises different stages, from the expression of the analytical need to the abandonment of the method, including method development, validation, selection, verification and routine use. Figure 10.1, adapted from Feinberg (2009), schematically illustrates the life cycle of an analytical method, and can be used to provide the context for the following paragraphs.

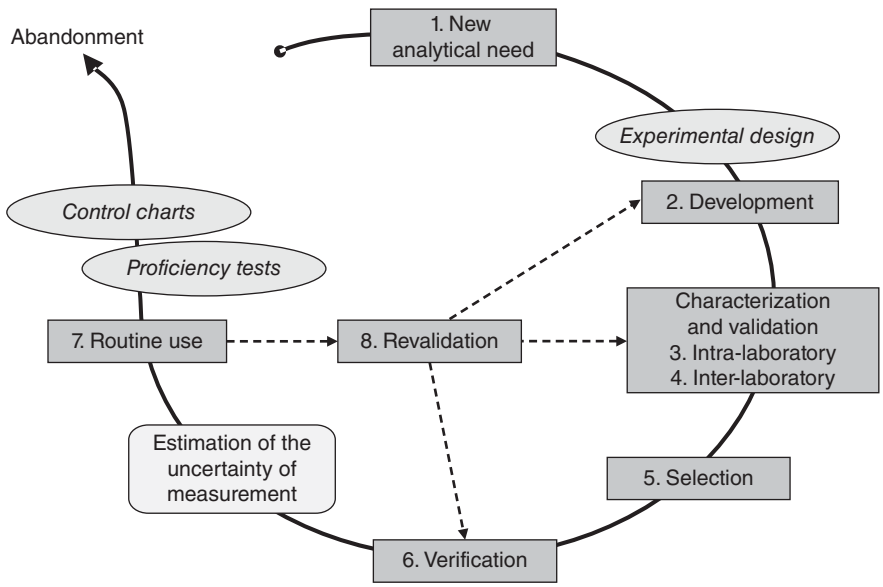


Fig. 10.1. Life cycle of analytical methods. The stages of development, characterization and validation are not necessarily implemented in the laboratory that wants to use the test routinely, and could have been performed by others. In that case, the laboratory 'only' has to implement the stages of selection and verification before routine use of the test.

10.3 Selection of molecular tests

Until recently, over and above the key aspects of a quality assurance system, such as standardization and formalization of procedures through to the production of SOPs (standard operating procedures) and staff training, a largely overlooked aspect of plant diagnostics was the effective choice of a test for a specific use and its subsequent characterization. Ensuring that the application of a test is 'fit for purpose' to detect the target pest within the desired host matrix, and knowledge of how the test performs against various criteria (such as sensitivity, specificity, repeatability and reproducibility), are essential for confidence in any result and for any decision to be taken on the basis of this result.

The choice of which test is to be applied in a given situation will be influenced by the sensitivity, speed, throughput and cost of testing, which should be driven by client demand (EPPO, 2010). For instance, testing asymptomatic plant material for a low-titre pathogen such as a virus will require greater sensitivity than the identification of an isolated pest to species level. Similarly, a test that is to be used in large-scale monitoring activities should be more amenable to high throughput and cost less to perform than a method designed for testing single samples. This selection step is often implicit, but it requires that validation data are available to enable a choice to be made among the tests that are available for the intended use.

10.4 Validation of molecular tests

10.4.1 Validation and verification of molecular tests

Tests used in a molecular laboratory must be evaluated and tested to define how reliable the results produced are for their intended purpose, i.e. they must be validated. Validation is defined in the standard ISO/IEC 17025:2005 (ISO/IEC, 2005) as 'the confirmation by an examination and a provision of objective evidence that particular requirements for a specific intended use are fulfilled'. Method validation is a distinct phase from method development and is performed after the method development is complete.

It is commonly agreed in plant pest diagnostics that a test is considered fully validated when it provides data for the following performance characteristics: sensitivity, specificity, repeatability and reproducibility (EPPO, 2014). If values for the test performance characteristics are not available or accessible, then the laboratory should produce the missing data and, consequently, must revalidate the test. When values for performance characteristics are available, there is no need to revalidate the test fully; the laboratory should 'only' verify that it can perform the test at the same level of performance (EPPO, 2014).

Verification is usually defined as a set of experiments conducted by a different analyst or laboratory on a previously validated test in order to demonstrate that in their hands the performance standards established by the original validation are attained. Verification is applicable only for standard or accepted methods that have been validated before. Usually, verification involves determining fewer characteristics and making fewer measurements for each characteristic than were determined in the original validation.

A rigorous validation process includes both single laboratory validation and collaborative study. A collaborative study (also called an inter-laboratory test performance study) is considered to provide a more reliable indicator of the performance of a test that is

to be used in other laboratories because it requires testing of the method in multiple laboratories, by different analysts, using different reagents, supplies and equipment, and working in different laboratory environments. However, even collaborative study for the validation of a test does not provide a guarantee of test performance in any laboratory that performs the test. This is why verification is essential.

10.4.2 Performance characteristics and the validation process

With the publication of the EPPO (European and Mediterranean Plant Protection Organization) standard PM7/98(2) (EPPO, 2014), laboratories working in plant pest diagnostics now have a clear framework to follow for the validation of a test, especially for molecular tests. A test, as defined by EPPO (2010), is 'the application of a method to a specific pest and a specific matrix'. It is important to note that in this case 'pest' refers to 'any plant, animal or pathogenic agent injurious to plants or plant products', as defined by the IPPC (International Plant Protection Convention) (IPPC, 2008).

An important point to remember is that the ultimate aim of validation is not to have the perfect test, but to understand the limits of any test in order to aid interpretation of the results. Within the EPPO diagnostic standards, the following assessment criteria are described (adapted from EPPO, 2010, and EPPO, 2014):

- Sensitivity: the lowest concentration of the target pest of which the presence can still be reliably established. The sensitivity is characterized through the establishment of the limit of detection (LOD). This is determined by carrying out a dilution series, or a range of concentrations, which should be relevant to the matrix/pest combination. For example, if the assay is to be used for the detection of a pathogen in macerated leaf material, the dilution series should ideally be made by diluting infected sap in healthy sap. A simple dilution of infected sap in buffer may give improved sensitivity, but does not mimic the application of the test. As a general guide, the recommendation is to test three dilution series or levels of concentration within a prescribed range of starting concentrations. This approach can be challenging for plant pathogens, as the starting concentration is unknown and a true analytical sensitivity cannot be determined, only a relative sensitivity, such as with plant viruses and phytoplasmas. Additionally, the LOD dilution series should be carried out a minimum of three times to verify the reliability of results, but if a reliable result is not obtained, further dilution series should be carried out until a reliable relative LOD can be established.
- Specificity: the ability of a test to distinguish the target pest from other organisms, whether related or not. Pathogens may occur in mixed infections and, wherever possible, it is highly recommended that cross-reactions from closely related species are ruled out. It is also possible that the test could cross-react with an unrelated species found on the host, and this consideration should also be included in the 'specificity' aspect of validation testing. Evaluation of the specificity of molecular tests often includes '*in silico*' experiments, e.g. checking molecular primers or probe sequences against known published sequences logged on a reliable sequence database, such as GenBank (the US National Institutes of Health genetic sequence database available from NCBI, the US National Center for Biotechnology Information). This *in silico* step must be completed by *in vitro* experiments, e.g. by performing molecular tests with biological material.

  - It is also important to determine the ability of a test to pick up as broad or narrow a range of strains/isolates as is required, in other words, to determine the analytical range of the test. For instance, it may be preferable to detect a pest or pathogen at the genus or group level, e.g. phytoplasmas, or an assay may be designed to be strain or biotype specific. To measure this, isolates should be tested from a range of strain types/populations and geographical areas to determine the effect of such variation on the test performance.

- Selectivity: an assessment of the relative insensitivity of a test to variations of the sample material. Even within a test matrix, such as the same host species, and with a specified material type (e.g. seed, leaf, bark, type of soil, etc.), there will be variation in test material. Physiological factors such as age of the host plant and variation between cultivars will affect the host matrix, so the performance of the test must be assessed under such variable conditions. Validation samples should account for an appropriate range of variation (e.g. cultivars, type of soil, etc.). If the test matrix is an isolated insect or nematode, these should be assessed through different life stages/instars.
- Repeatability: assessment of concurrence between assessments of the same measure under the same conditions. This is a measure of the sensitivity of the test to any variations that may occur under routine circumstances, and the assessment can be made concurrently with other validation assessments. For example, repeatability should be assessed at both medium and low concentrations, as this is where most variation in detection is likely to be seen. This evaluation could be carried out during the assessment of sensitivity.
- Reproducibility: concurrence between assessments of the same measure under varying conditions. The test should be run on different equipment, e.g. using different thermocyclers, on different days, operated by different people, but with a set of samples including the concentration/levels of dilution used for repeatability. This measure may also be carried out as part of a collaborative study.

Where possible, the validation process should be done using naturally infected material to measure how the test will perform during routine usage. However, in many cases, this may result in validation material that is inconsistent, as the pest (especially pathogens) may be unevenly distributed in the host matrix, e.g. as in natural seed infections. Artificially inoculated material may also be used, and may provide validation material that is both more reliable in the concentration and distribution and of a known strain, isolate, population or biotype. In practice, thorough validation should encompass all these sources of biological material. Basic evaluation of assay sensitivity, and assessments of repeatability and reproducibility, may be conducted with artificially produced samples, but specificity, repeatability and reproducibility should all encompass the testing of naturally infected samples so as to assess the performance of the test under realistic conditions.

10.4.3 Revalidation

In some cases, after a certain period of routine use, it will become necessary to abandon the test and to begin a new method life cycle (cf. Fig. 10.1) because the test has become

deprecated. In other cases, it could become necessary to make changes and, according to the importance of those changes, to revalidate the test (e.g. possible new intended uses on new matrices, or the description of new confusing species). If the change is minor, i.e. it concerns only a change in a particular operation during the test (such as 15 min in the water bath set at 65°C vs 20 min in the water bath set at 65°C), a judgement as to whether such a change requires complete validation or verification should be made and documented. If the change is significant, i.e. it affects the principle of the method (e.g. the type of extraction used), full revalidation is necessary. Whether the change is significant or minor will be decided by the technical manager, and in any case of doubt, the change should be considered as significant.

10.5 Implementation of a molecular test in terms of quality

10.5.1 Staff training

The implementation of quality assurance for a molecular test requires trained staff who can master the techniques that need to be used, and who are competent to perform them. Such staff need to understand the meaning of their work and actively contribute to the continuous quality improvement of the testing performed. For this reason, it is particularly important to define the responsibilities and functions of the different people involved in the implementation of the test and the validation of the test results. The two key functions are those of technical manager and technical operator. These two functions can be performed by the same person, but deputies must be appointed.

Technical managers have the technical expertise. They know how to implement the test, organize its implementation, technically validate the results and ensure technical and scientific monitoring. They also ensure the quality management relating to the test (quality document drafting and review, treatment of the feedback information, organization of the quality controls).

Technical operators are responsible for the implementation of the test in the laboratory. They prepare the buffers, implement the analytical protocol in accordance with the quality management system, and ensure the cleaning and the decontamination of equipment and materials used for the analyses.

It is essential to define the minimum levels of qualification and experience necessary for these key functions and to ensure that those requirements are met on the basis of appropriate education, training, experience and/or demonstrated skills. In addition, a competence evaluation of the technical manager and the technical operator must be performed regularly.

When the training process is done correctly, there should be no dilution of skills or knowledge, which in turn should ensure that standards are maintained. Generally, training is broken down into four stages, in which the demands are increased with the increasing competence of the trainee. At the completion of each stage of training,

the trainee should be certified as competent to move on to the next stage of training. This is normally done by the trainer or other competent person signing the trainee's training record. The four stages of the process are:

- Familiarization (Stage 1): in this stage, which is often called 'Preparation', staff should familiarize themselves with the SOPs in which they are being trained. This should include reading both the SOP(s) and any accompanying documentation, such as chemical hazard sheets and risk assessments. This should not mean that the trainee is left to wade through an SOP without guidance from the trainer, as the purpose of this stage is to give a clear understanding of the theory behind the process the trainee is learning.
- Demonstration (Stage 2): the trainee will shadow the trainer through the process. At this stage, the trainee should also be following the SOP so that he or she can visualize the parts of the process that were read about in Stage 1. The trainer should explain the correct way to carry out a process and why these points are important. The trainer should also take care to point out common sources of error so that the trainee becomes aware of where/when things may go wrong. In many cases, it is appropriate that this aspect of the training is carried out more than once.
- Initial Competence Assessment (Stage 3): the trainee executes the process under supervision from the trainer. This is to ensure that the process is carried out and accomplished, to the SOP according and that the trainee demonstrates the necessary skills required to achieve consistent results. For this reason, this stage of training requires the trainee to complete the process successfully on multiple attempts. In the early attempts, the trainer may give guidance while watching the trainee, but over successive occasions of running through the process, the trainee should complete the process unaided.
- Competence Assessment (Stage 4): after the trainees have been signed off as successfully completing the performance trial, they are considered to be competent in carrying out the process. The trainees then enter an ongoing cycle of competence assessment to ensure they continue to perform to the required standards. Where a process is performed on a regular basis, this assessment can be done by monitoring the test results of the individual over time to ensure that the results are consistently within acceptable parameters. Alternatively, where a process is carried out infrequently, the trainee may be required to complete a performance trial on mock samples to ensure that he or she can perform to the accepted standard.

10.6 Infrastructure and equipment

Access to the laboratory should be restricted to authorized personnel only, who should be aware of the intended use of a particular area and the restrictions imposed on working in such areas. The organization of facilities and equipment must allow the progress of the test samples and minimize the risks of cross-contamination. Detailed guidance on the implementation of quality assurance for accommodation and equipment is given in EPPO PM7/84 (EPPO, 2007). According to the AFNOR (Association Française de Normalisation) standard XP V03-043:2008 (AFNOR, 2008), the laboratory facilities

dedicated to molecular biology should feature at least three (b, c, d below), but preferably four, zones that are physically segregated from each other (microbiological safety hoods, different rooms, etc.) (AFNOR, 2008):

(a) one zone for the preparation of test samples;

(b) one zone for nucleic acid extraction;

(c) one zone for the pre-amplification steps (preparation of master mixes, addition of nucleic acids); and

(d) one zone for the post-amplification steps (opening of tubes and handling of amplification products).

Workspaces must be sufficient, appropriate to the activities to be carried out, regularly cleaned and appropriately decontaminated between different samples and/or activities. All gloves (which should be non-powdered to avoid the risk of inhibition and/or contamination involved with PCR reaction mixtures) and laboratory coats should be changed on an appropriate rotation.

The laboratory should also monitor, control and record environmental conditions that may influence the quality and reliability of the molecular tests. For example, the impact of temperature, and particularly of room temperature, on molecular test results is well known. Thus, the temperature should be controlled in at least the pre-amplification and the post-amplification zones. The pressure should also be set between these two zones to avoid the risk of backtracking of materials, e.g. aerosols. It is also recommended that dedicated laboratory coats are used in the pre-amplification and post-amplification zones.

As a further measure to minimize the risk of cross-contamination, it is preferable that the equipment is dedicated to a specific zone and activity. In particular, each zone of activity should have its own set of micropipettes (with hydrophobic filter tips to avoid aerosol–vector cross-contamination).

The equipment considered to be critical for the implementation of the test (i.e. that which significantly affects the test results) must be controlled. This equipment (micropipettes, balances, thermostatic chambers, thermocyclers) should be listed and identified, and a programme should be documented and implemented for the maintenance and metrological control of the critical pieces of equipment. In particular, the quantitative thermocyclers composed of an optical system and the apparatus for reading fluorescent-labelled gel electrophoresis patterns must be subject to regular maintenance according to the manufacturer's recommendations.

10.6.1 Instrument calibration

The metrological operations performed should provide traceability to the international system (ISO/IEC, 2005) when available. The control of the thermocyclers, which is specific to molecular tests, needs to be explained. It is recommended that a thermocycler is 'qualified' at least once a year, and also after each maintenance operation. Two types

of procedures can be used for qualifying a thermocycler: (i) a biological test and (ii) metrological verification. It is possible to use one or both of these procedures because they are complementary.

In the biological test, the molecular test if performed using adequate biological controls to ensure the optimum performance of the thermocycler under routine conditions. The test must be performed:

- for each well or capillary used in routine testing (or a sample of wells or capillaries representative of the functioning and the use of the apparatus);
- for each hybridization temperature (or each temperature range);
- for each operating mode (fast vs normal) for the relevant apparatus; and
- using samples at a concentration close to the detection limit.

If one or several wells do not amplify during the biological test, the test must be performed again for the wells concerned and also for the adjacent wells. If amplification is obtained for all of the wells, the apparatus is qualified; if not, the laboratory must perform a cause analysis and implement curative and corrective actions (e.g. cleaning the apparatus, sending the apparatus for repairs, eliminating the defective wells from the routine analyses, etc.). The biological test ensures a monitoring of the overall performance of the method (including the use of the thermocycler), but not specifically the performance of the thermocycler. It does not allow the detection of possible drifts in temperature that may affect the specificity of the methods.

The metrological verification of a thermocycler consists of controlling the trueness and the homogeneity of thermal profiles from a sample representative of the wells (or capillaries) and cycles (AFNOR, 2015). There are different PCR thermal cycler calibration devices and methods. The device used must be appropriate to the thermocycler to be tested and must ensure adequate representation, such as that defined by the manufacturer. The values obtained during the trueness and homogeneity tests should not differ from the requirements defined by the laboratory for the implementation of PCR analyses (generally in the range of ±1°C from the specified temperature for trueness and in the range of ±2°C for homogeneity). If a temperature deviation is observed (e.g. an overshoot or undershoot), the laboratory must evaluate the consequences, or the absence of consequences, of extreme temperature profiles on the results of the analyses. In this situation, the laboratory can, for example, use appropriate controls and/or implement a biological test and/or retest the analytical specificity of the method.

10.7 Critical reagents

Reagents used for molecular tests must be analytical grade reagents adapted for use in molecular biology applications, and must be stored and used under the appropriate conditions (AFNOR, 2008).

Critical reagents (i.e. those that significantly affect the test results) for extraction and amplification (e.g. the extraction kit, *Taq* DNA polymerase, primers, probes, deoxynucleotide triphosphates – dNTPs) must be checked:

- either before their routine use, in which case new batches of reagents are validated in comparison with the old batches, or in comparison with reference criteria using reference material (determined during the validation of the test);
- or by a posteriori control. In this case, the use of appropriate test controls (cf. Section 10.10.1, Internal quality controls, Test controls) allows for validation of the new batches used. Test controls also allow the recording and monitoring of reagent quality during their period of use.

10.8 Measurement uncertainty

ISO/IEC Guide 99:2007 (ISO/IEC, 2007) defines the uncertainty of measurement as a non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurement based on the information used. An estimation (or at least a full consideration) of the components contributing to the overall uncertainty of a measurement (or a test result) provides a means of establishing that the results obtained are valid. It also confirms that the item under test is fit for the intended purpose.

Estimation of the uncertainty of measurement is generally carried out in two stages:

1. The first stage is the identification of the sources of uncertainties.
2. The second stage is the quantification of uncertainties. The total uncertainty of a measurement is found by combining all of the contributing component uncertainties.

The nature of the analytical methods used for molecular testing in plant health (mainly qualitative methods) does not allow rigorous calculations (in terms of metrology and statistics) to evaluate the uncertainty of measurement. In particular, the quantification step is difficult to implement. Thus, for each test, it is recommended that the sources of uncertainties (or critical points) are identified, i.e. that the steps of the analytical process that pose risks which may influence the analytical results are identified. This is done in order to implement appropriate measures to reduce and control the risk at such a level that its contribution to the overall uncertainty can be considered as being negligible.

Identification of the sources of uncertainties begins by examining the measurement process in detail. The Ishikawa diagram (also known as the fishbone diagram, the cause and effect diagram or the 5M diagram) can be used to identify the sources of uncertainties (factors of influence/critical points) (Feinberg, 2009). This minimizes the risk of leaving out some of the sources of uncertainties by sorting them into large families:

- material – raw material, quality, supply, etc.;
- machine – machinery, tools, equipment, maintenance, metrology, etc.;
- manpower – direct, indirect, experience, competence, etc.;

- milieu (environment) – physical environment, dust, location, layout, temperature, legislation, etc.; and
- method – instructions, procedures, modus operandi, etc.

If some factors have a contribution that is considered to be insignificant, they can be overlooked. An example of a 5M diagram applied to the evaluation of the uncertainty of measurement for molecular tests is given in Fig. 10.2. EPPO PM7/98(2) (EPPO, 2014) presents real examples of laboratory reports on the critical points relating to uncertainty of measurement in the implementation of molecular tests. This approach, which is based on risk analysis, is of interest in demonstrating control of the analytical process.

10.9 Documentation

As with any laboratory, the quality management system of a molecular laboratory must be based on a controlled and up-to-date documentation system (ISO/IEC, 2005). This documentation system consists of declaration documents (quality manual, possibly quality plans), explanatory documents (SOPs, instructions) and probative documents (records). Declaration documents describe the policy and the objectives of the quality management system. Explanatory documents describe the way to carry out specific

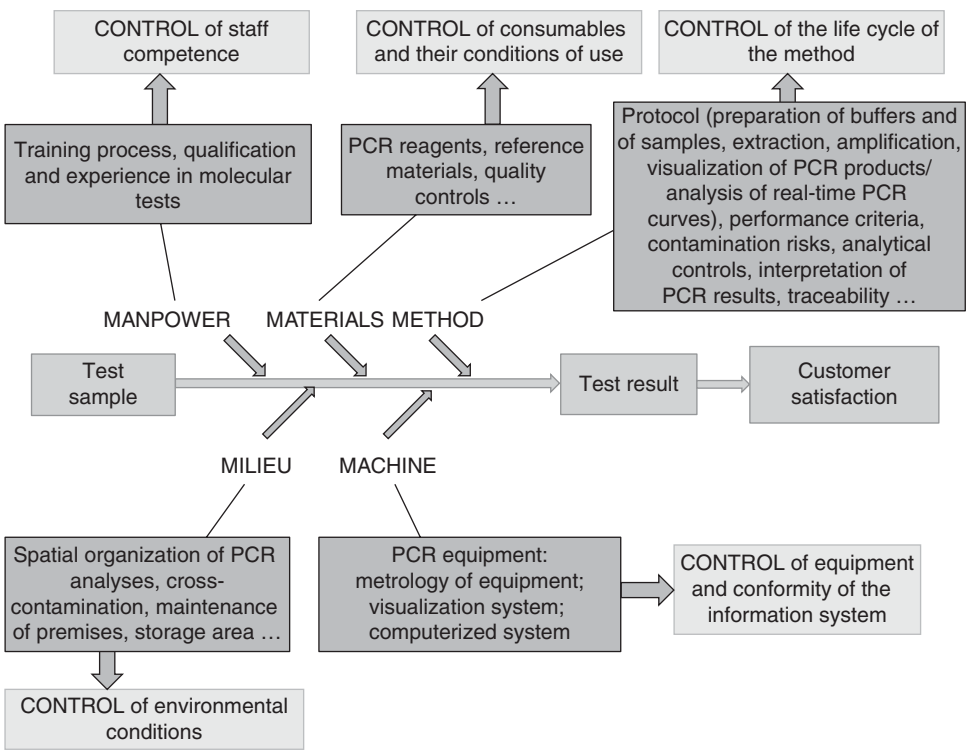


Fig. 10.2. 5M diagram applied to the evaluation of uncertainty of measurement for molecular tests.

tasks. The records ensure the full registration of the information associated with the achievement of a task; they include forms, worksheets, check sheets, test reports, calibration certificates, customers' feedback and other forms. All of these records provide traceability for molecular tests, which is particularly important in case of a problem (non-conformity, deviation), so that, for example, the analyses affected can be identified and those that must be repeated specified. If done effectively, the laboratory recording systems should provide a network of data linking who carried out which task and their competence for the task, as well as which machine was used, and whether it was appropriately calibrated and maintained. Tracking and monitoring trends in performance and, most crucially, acting upon those data, allow continuous improvement of the activity.

10.10 Survey of test performance and quality management

10.10.1 Quality controls

During the routine use of a test, the laboratory should implement a system of controls to ensure the reliability of the results and monitor the efficacy of the test procedures on an ongoing basis.

Two types of controls can be implemented:

- internal quality controls, e.g.
  - the effective use of test controls made from material considered as a reference;
  - the use of control charts (monitoring of the analytical process and the identification of trends);
- external quality controls, such as participation in inter-laboratory comparison or proficiency-testing programmes, when they exist.

Internal quality controls

Test controls At the most basic level, the use of test controls gives a measure that a test is working correctly in detecting the appropriate target pest (a positive control), and not giving positive reactions to similar sample material (a negative control). This approach has been the basis of the determination of results in microplate ELISA testing since the first description of this method by Clark and Adams (1977). However, with the advent of molecular methods, the advantage of greater sensitivity also gave an increased risk of cross-contamination during nucleic acid extraction and PCR preparation, and resulted in the need for further controls. Table 10.1 provides an overview of the controls that are recommended within standard molecular testing protocols, such as those published by EPPO and other standardization bodies (e.g. IPPC).

Before any interpretation can be run on the samples tested, it is essential first to validate the handling procedure carried out on the series of samples. A handling procedure is validated only if the results obtained from the controls used at each step of the handling

Table 10.1. Controls to be used in molecular testing. (Adapted from EPPO, 2013, and XP V03:2008 (AFNOR, 2008).)

Control type	Description	Purpose	Frequency
Negative process control (NPC)	A reference sample not contaminated or infested with target organism that should be put through the same process steps as the test samples	To detect the presence of cross-contamination during the nucleic acid extraction phase	An NPC should be processed with every batch of samples. Where batches of samples are large, more than one NPC can be included and should be dispersed within the batch and be treated as mock samples
Negative amplification control (NAC)	A PCR reaction performed with molecular grade water (RNA and DNA free) and without any PCR inhibitor	This control should detect false positives arising from contamination occurring during preparation of the reaction master mix or during amplification	NACs should occupy a minimum of 2 wells in every master mix on a plate. If issues do occur with contamination due to aerosols during spiking, these controls should be placed with increased frequency across the plate
Positive process control (PPC)	A well-characterized reference sample contaminated or infested with a detectable amount of the organism that undergoes nucleic acid extraction and amplification in parallel with test samples. If naturally infected host/sample matrix is not available, the target organism may be artificially added to the host/sample matrix before extraction. The PPC goes through exactly the same process steps as the test samples	This control is a measure of the quality of extraction for the target nucleic acid. In the case of quantitative methods, it may also be a measure of the amount of target nucleic acid recovered. The control also allows possible inhibition to be assessed during amplification. Ideally, these controls should include a high-level positive as well as a limit of detection (LOD) positive to ensure that there is no drop-off in sensitivity	A PPC should be included in every batch of samples processed. Ideally, these should be processed at the end of the batch to limit cross-contamination

Continued

Table 10.1. Continued.

Control type	Description	Purpose	Frequency
Positive amplification control (PAC)	A PCR reaction performed with a solution containing a detectable amount or copy number of the target DNA sequence	A measure of the efficiency of the amplification reaction. Ideally, this control should be close to the LOD determined during the validation process. When close to the LOD, this control is used to demonstrate that the PCR reaction has been optimized	A PAC should be included on every plate for all tests. Where testing of samples crosses more than one plate, the same PAC should be used on each plate
Inhibition control (IC)	There are two possible types of inhibition control, an: • **internal inhibition control**, in which the amplicon is acting as an internal control, and is obtained during the amplification reaction of the target fragment by adding DNA and/or primers. This amplicon is clearly different from the target fragment; • **external inhibition control**, in which the amplicon is acting as an external control, and is obtained through a separate amplification reaction from that of the target fragment (i.e. in a different reaction tube). This amplicon may or may not be different from the target fragment	This control enables the laboratory to check that there has been no PCR inhibition affecting the results. An inhibition control makes it possible to interpret a negative result unambiguously, as it highlights the false negatives obtained in the presence of PCR inhibitors	All samples should be tested for inhibition control

procedure conform to the expected results. The individual results obtained from the samples tested can then be scored according to the requirements specified for each test.

If a control does not provide the expected result, the information must be used to identify the cause of the problem. For example, if the positive process control provides a negative result, but the positive amplification control is positive, a problem during the extraction of nucleic acids should be suspected and investigated.

It is worth noting that the use of controls allows a global validation of the analytical process, in particular that the consumables and reagents used for the test are appropriate and of good quality, that the operator has correctly implemented the test, etc.

Use of reference materials A reference material is defined in ISO GUIDE 30:2015 (ISO, 2015) as 'a material, sufficiently homogeneous and stable with respect to one or more specified properties, which has been established to be fit for its intended use in a measurement process'. Reference materials provide essential traceability in tests and should be used to produce the working material used to perform method validation and verification, but also to demonstrate the accuracy of routine results (by the use of test controls). For molecular tests, reference materials can consist of live cultures, infected matrix or plant material; they can also be DNA/RNA preparations. Preferably, material should come from a recognized source, such as a national reference collection (e.g. reference cultures) or specialized providers (commercial controls). Laboratories may also produce their own 'reference' materials. Before the reference material is used (for validation purposes or in routine analyses), it should first be ensured that the material is producing the features for which it was selected, and then regular verifications should be carried out to control its integrity in the laboratory's conditions of storage and use.

Control charts A control chart is a graphic representation of process variability. It is used to monitor the stability of a process (e.g. a molecular test) over time (ISO, 2014).

Two types of variations can occur in any process, depending on the cause, which may be random or special. A process is considered to be in statistical control if it is affected by random causes only (which are inherent in the process), i.e. if no special causes (non-random causes) have entered the system. Such special causes can affect the level at which the process is operating, the degree of variability around the process level, or both simultaneously. In a molecular test, special causes can affect the test results. So, it is important to identify and try to eliminate special-cause variation. Out-of-control points and non-random patterns on a control chart indicate the presence of special-cause variation.

The Shewhart control chart (ISO, 2013), which is the most common type of control chart, can be used for monitoring real-time PCR tests where quantitative data – cycle threshold (Ct) values – are available. This involves producing a batch of sufficiently homogeneous and stable positive amplification control (PAC) material. This PAC will be run with every batch of test samples and the Ct value obtained will be plotted on the chart. The chart requires starting with at least 20–30 values to calculate a mean and a

standard deviation(s), which form the basis for control values equivalent to the mean ±2 s (warning limits) and the mean ±3 s (control limits). If the analytical process is 'in statistical control', not more than 5% of the values fall in the warning zone. Any value falling above the rejection limit, or two consecutive values in the warning region, suggest the introduction of a new (and most likely unanticipated) source of variation, which is known as a special-cause variation. The presence of a special-cause requires investigation and corrective action.

The main requirement for this type of control chart is that the data must be normally distributed. Consequently, it is preferable to choose the PAC with a medium level of contamination (of 5–100 times the LOD). Figure 10.3 provides an example of a control chart used to monitor a real-time PCR test for the detection of a pathogenic bacterium (*Xanthomonas axonopodis* pv. *allii*) of onion.

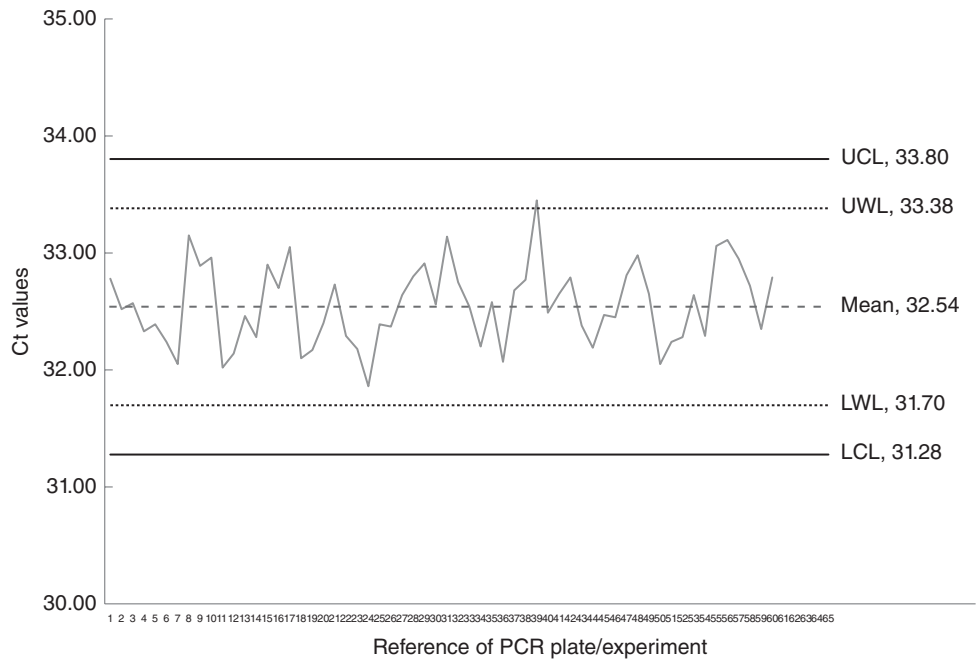


Fig. 10.3. Example of a control chart used in plant health: a Shewart X-bar control chart for a real-time PCR test (detection of *Xanthomonas axonopodis* pv. *allii* in onion seed). The test control plotted on the chart is a positive amplification control, which consists of a bacterial solution containing about 5×10^3 (= 5000) target cells/ml stored at a temperature lower or equal to −18°C in single-use aliquots. This control is included on each plate for each experiment, and each Ct (cycle threshold) value obtained for this control (one per plate) is plotted on the chart (middle grey line). The reference values associated with this test control (mean and standard deviation(s)) were established from 30 experiments performed under different test conditions (operators, days, etc.) and are used to determine: (i) the target Ct value (grey dashes; called 'mean' on the chart) for the test control; (ii) the control limits (solid black lines: UCL, upper control limit, mean +3 s; and LCL, lower control limit, mean −3 s); and (iii) the warning limits (dotted black lines: UWL, upper warning limit, mean +2 s; and LWL, lower warning limit, mean −2 s).

External quality controls

Participation in proficiency tests (PTs) is a highly effective way to validate the laboratory's analytical process. Indeed, such tests are designed to assess and demonstrate the ability of participating laboratories to implement assays or measurements. They also constitute, for the participants, a genuine tool for improvement, as long as the examination of results is correctly performed to identify and trigger corrective actions, if these are needed. Participation in PTs complements the internal quality control procedures by providing an additional external assessment of the competence of testing laboratories. It also provides additional confidence to laboratory clients, as well as for other interested parties.

The laboratory should plan to participate in sufficient proficiency-testing activities to cover the diversity of its molecular testing: for example, type of technique; type of extraction (kit, immunocapture, etc.); type of PCR (conventional PCR, real-time PCR); type of matrix (plant, soil, water …); type of pest (bacterium, virus, fungus, nematode, insect …). A well-constructed PT plan will allow the laboratory to meet quality requirements and verify the analytical processes associated with the laboratory scope in molecular biology at minimum expense.

The increasing development and use of proficiency tests has led to the development of accreditation for PT providers, according to ISO/IEC 17043:2010 (ISO/IEC, 2010). Participation in PT organized by accredited providers provides the participants with a reasonable degree of confidence in their decision-making process, based on the results from the proficiency testing (Chabirand *et al.*, 2014).

10.11 Continuous improvement

Any deviation from the norm or non-conforming work during the implementation of the molecular test must be recorded. The consequences of this deviation for the test results must be evaluated by competent personnel. A cause analysis must be conducted and corrective actions must be implemented to prevent the problem from recurring. This allows the laboratory to improve its management system continually, including testing.

Other tools can also be used to improve the quality of the test:

- the treatment of client complaints;
- the non-conformities and deviations reported in audits;
- the implementation of preventive actions;
- the analysis of data and trends (in particular, the analysis of indicators, quality control results);
- the process review and management review; and
- the collection of client feedback (e.g. through satisfaction surveys).

10.12 Conclusion

The implementation of a quality management system in a molecular laboratory performing tests in plant health contributes to improvement in the reliability of molecular test results, which is essential in order to guarantee plant health and the control of plant diseases or pests. It also can be used as a competitive advantage in the framework of trade. Indeed, a laboratory that can produce test results at the same cost as its competitors, but is accredited/recognized (or has this in process), will have an advantage over those laboratories that are not so cost-effective or accredited/recognized.

Quality management should not be seen as cumbersome, binding or costly. On the contrary, it should serve as a real tool for management and performance improvement. In this sense, an approach based on risk management may be helpful. According to ISO 31000:2009 (ISO, 2009), risk is 'the effect of uncertainty on objectives', whereas risk management is 'coordinated activities to direct and control an organization with regard to risk'. Risk management is not a new concept, just a formal description of the activities that laboratories are already performing as part of their quality assurance programme for preventing errors and reducing harm to the test results. Risk management helps laboratories to find the optimal balance between their quality control strategy and the relative risks.

References

AFNOR (2008) *XP V03-043:2008, General Requirements for Molecular Biology Analysis for Detection and Identification of Pathogenic and Destructive Organisms in Plants and Derived Products*. AFNOR Groupe, La Plaine Saint-Denis, France.

AFNOR (2015) *NF U47-600-1:2015, Animal Health Analysis Methods – PCR – Part 1: Requirements and Recommendations for the Implementation of Veterinary PCR*. AFNOR Groupe, La Plaine Saint-Denis, France.

Chabirand, A., Anthoine, G., Pierson, O. and Hostachy, B. (2014) The organization of proficiency testing in plant pathology (qualitative methods of analysis) according to the ISO/IEC 17043: example of the French national reference laboratory. *Accreditation and Quality Assurance* 19, 111–125.

Clark, M.F. and Adams, A. (1977) Characteristics of the microplate method of enzyme-linked immunosorbent assay for the detection of plant viruses. *Journal of General Virology* 34, 475–483.

EPPO (2007) PM7/84. Basic requirements for quality management in plant pest diagnosis laboratories. *EPPO Bulletin* 37, 580–588.

EPPO (2010) PM7/76(2). Use of EPPO diagnostic protocols. *EPPO Bulletin* 40, 350–352.

EPPO (2013) PM7/113(1). *Pepino mosaic virus*. *EPPO Bulletin* 43, 94–104.

EPPO (2014) PM7/98(2). Specific requirements for laboratories preparing accreditation for a plant pest diagnostic activity. *EPPO Bulletin* 44, 117–147.

Feinberg, M. (2009) *Labo-Stat: Guide de Validation des Méthodes d'Analyse*. Editions Tec et Doc Lavoisier, Paris.

IPPC (2008) *ISPM (International Standards for Phytosanitary Measures) No. 5 Glossary of Phytosanitary Terms*. Secretariat of the International Plant Protection Convention, Food and Agriculture Organization of the United Nations, Rome.

ISO (2009) *ISO 31000:2009, Risk Management – Principles and Guidelines*. International Organization for Standardization, Geneva, Switzerland.

ISO (2013) *ISO 7870-2:2013, Control Charts – Part 2: Shewhart Control Charts*. International Organization for Standardization, Geneva, Switzerland.

ISO (2014) *ISO 7870-1:2014, Control Charts – Part 1: General Guidelines*. International Organization for Standardization, Geneva, Switzerland.

ISO (2015) *ISO GUIDE 30:2015, Reference Materials - Selected Terms and Definitions*. International Organization for Standardization, Geneva, Switzerland.

ISO/IEC (2005) *ISO 17025:2005, General Requirements for the Competence of Testing and Calibration Laboratories*. International Organization for Standardization, Geneva, Switzerland.

ISO/IEC (2007) *ISO/IEC GUIDE 99:2007, International Vocabulary of Metrology – Basic and General Concepts and Associated Terms (VIM)*. International Organization for Standardization, Geneva, Switzerland.

ISO/IEC (2010) *ISO 17043:2010, General Requirements for Proficiency Testing*. International Organization for Standardization, Geneva, Switzerland.

Index

Note: bold page numbers indicate tables and figures. Numbers in brackets preceded by n are chapter endnote numbers.