Research Review No. 91140002-03

Soil Biology and Soil Health Partnership Project 3: Molecular approaches for routine soil-borne disease and soil health assessment – establishing the scope

John Elphinstone¹, Bryan Griffiths², Matthew Goddard³ and Elizabeth Stockdale⁴

¹ Fera Science Ltd., Sand Hutton, York, YO411LZ;
² SAC Commercial Ltd., SRUC Edinburgh Campus, King's Buildings, West Mains Road, Edinburgh, EH93JG;
³ University of Lincoln, Brayford Pool, Lincoln, LN67TS;
⁴ NIAB, Huntington Road, Cambridge, CB30LE

This review was produced as the final report of a 6 month project (Project 3) within the Soil Biology and Soil Health Partnership (AHDB: 91140002) which started in June 2017. The work was funded by a contract for £9996 from AHDB and BBRO.

While the Agriculture and Horticulture Development Board seeks to ensure that the information contained within this document is accurate at the time of printing, no warranty is given in respect thereof and, to the maximum extent permitted by law, the Agriculture and Horticulture Development Board accepts no liability for loss, damage or injury howsoever caused (including that caused by negligence) or suffered directly or indirectly in relation to information and opinions contained in or omitted from this document.

Reference herein to trade names and proprietary products without stating that they are protected does not imply that they may be regarded as unprotected and thus free for general use. No endorsement of named products is intended, nor is any criticism implied of other alternative, but unnamed, products.
CONTENTS

1. ABSTRACT .............................................................................................................................. 1

2. REVIEW OF MOLECULAR APPROACHES TO BIOLOGICAL SOIL HEALTH ASSESSMENT .......................................................................................................................... 3

   2.1. Introduction .......................................................................................................................... 3

   2.2. Sampling ............................................................................................................................... 3

   2.3. DNA extraction and purification .......................................................................................... 4

   2.4. Targeted analyses using quantitative polymerase chain reaction (qPCR) ......................... 5

   2.5. Methods suitable for soil community analyses ................................................................. 7

       2.5.1. Microarrays .................................................................................................................. 7

       2.5.2. DNA fingerprinting methods ....................................................................................... 7

       2.5.3. Next generation high throughput sequencing ............................................................. 8

   2.6. Molecular markers for bio-indicators of soil health .......................................................... 10

       2.6.1. Biological communities contributing to soil biodiversity ............................................. 10

       2.6.2. Microbial activity affecting soil nutrient availability/retention .................................. 11

       2.6.3. Microbial symbionts contributing to crop nutrition .................................................... 14

       2.6.4. Biocontrol agents ......................................................................................................... 15

       2.6.5. Nematode assemblages ............................................................................................... 16

       2.6.6. Soil-borne plant pathogens ................................................................---------------------- 16

3. CONCLUSIONS ........................................................................................................................ 19

4. REFERENCES .......................................................................................................................... 21

5. APPENDIX 1 ............................................................................................................................ 33

6. APPENDIX 2 ............................................................................................................................ 34
1. Abstract

This review is part of a suite of integrated projects (Soil Biology and Soil Health Partnership) specifically aimed at addressing the AHDB and BBRO Soils Programme call – “Management for Soil Biology and Soil Health”. This project is designated Project 3 within WP 1 (Benchmarking and Baselining; see Figure 1).

This review specifically aimed to:

- Summarise available knowledge on procedures to sample field soils to undertake DNA analysis of the presence and composition of microbial communities and their functions to provide indicators of soil health.
- Evaluate standard tools for use in routine sampling and molecular analysis of soil quality so that their value can be demonstrated to growers and agronomists during and beyond the current Soil Biology and Soil Health Research Partnership.
- Establish full lists of molecular markers that can be used to quantify:
  (a) Soil-borne pathogens for use in prediction of crop disease;
  (b) Indicators of good soil health which can influence crop yield and value.

Procedures for sampling soil and extracting DNA from the sample have been reviewed. There are no standardised sampling methods, but it is general practice to take composite samples by mixing multiple cores from the surface to 10-30 cm depth. Corers should be cleaned and flamed between collection of each separate set of composite samples. A sample size of at least 200-500g is recommended. The number and spatial arrangement of samples and sub-samples required depends on the expected distribution of the target. For unknown target distributions, it has been suggested that the area is divided into evenly-sized grids with at least 2 composite samples per grid. An internationally recognised standard (ISO 11063:2012) describes a procedure for direct isolation of DNA from soil, suitable for further analysis using qPCR and high throughput sequencing methods, but this does not include procedures for subsequent DNA purification and is only suitable for small soil samples. Procedures more suitable for direct extraction and purification of total DNA from composite soil samples of 200-500g are described in the Appendices to this report.

Various molecular techniques have been used for analysis of soil quality, including methods based on polymerase chain reaction (PCR), microarrays, DNA fingerprinting (DGGE and T-RFLP) and DNA sequencing. Two approaches are considered most suitable for routine analysis of taxonomic or functional markers; quantitative PCR (qPCR) for detection and quantification of specific markers and next generation high throughput sequencing for analysis of whole soil communities. The range of molecular markers that have been used to investigate the taxonomy and function of individual target organisms and communities of organisms in soil are described in full. These include taxon-specific markers, mainly based on selected DNA sequences from within ribosomal (rDNA) or mitochondrial (cytochrome oxidase) DNA loci. Functional markers in genes expressing key enzymes involved in
carbon, nitrogen, phosphorus and sulphur cycling are also described. Markers are also listed that have been used to assess soils for presence and activity of other key bioindicators of soil health, including mycorrhizal fungi, nitrogen-fixing microorganisms, plant growth promoting bacteria, biocontrol agents, nematode assemblages and plant pathogens.

Some technical challenges remain to be fully overcome in the application of these technologies to ensure a representative and unbiased analysis of soil microbiological communities and their function. These include further standardisation of procedures for sampling, extracting and purifying DNA from soils, improved consistency in the choice of markers to be used in the analyses and the use of appropriate internal controls that ensure accuracy of data interpretation. The high cost of molecular analysis also remains a constraint to its routine application.

Figure 1. Diagram to show how Project 3 (in black) fits into the organisation of the Soil Biology and Soil Health Partnership.
2. Review of molecular approaches to biological soil health assessment

2.1. Introduction

In the last decade several useful biological indicators for soil health assessment have been proposed (Ritz et al., 2009). There have also been advances in knowledge of the functional roles of below-ground biodiversity (Bardgett & van der Putten, 2014) and major advances in technology and a reduction in the cost of using molecular tools (Orgiazzi et al., 2015). The following review of a rapidly developing literature aims to describe those molecular procedures that currently offer the most potential for detection and quantification of individual target organisms or functions, or that allow analysis of the diversity of whole communities of organisms or their functional groups. The current range of DNA markers available for analysis of the biological component of soil health is also reviewed. Whilst most approaches have so far been used in a research context, their suitability for application in affordable routine soil health analyses has yet to be established.

Appropriate procedures for sampling soil and extracting DNA from the sample are also reviewed. Sampling methods capable of recognising in-field variation and methods to extract high quality DNA from statistically representative soil samples are required. Extraction methods need to be efficient across all soil types and farming practices. They also need to be geared towards the type of molecular analysis to be performed, since some methods require higher levels of purification or longer DNA fragments than others. Methods also need to be suitable for direct extraction and analysis of total soil DNA, of DNA indirectly extracted from isolated soil organisms or for extraction of e-DNA (DNA remaining in the environment after release by organisms into the soil).

2.2. Sampling

To ensure that soil samples are representative of the area being sampled, and to account for heterogeneous distribution of soil micro- and macro-fauna, the sample is usually prepared as a composite of several small cores (20-50g each). Cores are typically 100 mm long and 10 mm diameter. The standard practice is to collect sub-samples randomly across the area to be sampled and to include material from different depths, usually topsoil only and down to 10-30 cm, depending on the expected range of distribution of the target organisms. The size of the study area should also represent the expected distribution of targets. To monitor variability in nematode distribution introduced by the sampling strategy, Taberlet et al. (2012) recommend sampling across a regular grid pattern across the study area with at least 2 randomly collected composite samples collected per grid. Corers should be suitably cleaned to remove soil remains followed by alcohol flaming between the set of soil cores that will together form one composite sample.
The size of each composite soil sample has been optimised for molecular test methods (Wiesel et al., 2015). Soil samples for nematode extraction have traditionally been standardised at 200g (Flegg and Hooper, 1970). Although a wide range of sample sizes between 1-200g have been employed in molecular studies (e.g. Waite et al., 2003; Porazinska et al., 2010; Griffiths et al., 2012; Morise et al., 2012), it was recently concluded that a sample of at least 200g is required for accurate measure of nematode abundance and at least 100g is needed to truly represent community composition (Wiesel et al., 2015). Similarly, for molecular study of soil-borne fungal and bacterial populations, Ophel-Keller et al. (2008) collected 3 samples per field (40-300 ha) with 15 cores per sample on a ‘W’ transect. They suggested that DNA should be extracted from samples of at least 200-500g for ‘biologically relevant analysis’. For molecular analysis of earthworm DNA in soil (Bienert et al., 2012), 2 layers (0-20cm and 20-40cm) have been sampled by randomly collecting subsamples to make a 500g sample per area of 10m radius.

2.3. DNA extraction and purification

Numerous procedures are described for direct extraction of total soil DNA or eDNA or indirect extraction of DNA from isolated soil organisms. The yield of extracted DNA varies with the chemical and physical properties of different soils (Feinstein et al., 2009), including the contents of negatively charged clay, silicates and organic matter which bind to DNA. The yield also varies according to the direct extraction process used and the relative efficiency of cell lysis of the different target organisms (Petric et al., 2011). The quality of extracted DNA can also vary with the method used; more intensive treatments may improve microbial cell breakdown resulting in higher DNA yields but may cause shearing of the DNA making it unsuitable for some types of molecular analyses. Different DNA extraction methods can therefore bias the results of molecular analyses in different ways. The choice of DNA extraction method therefore depends on the soil type, the target organism(s), the type and quality of DNA to be studied and the method being used for its analysis. The preferred extraction method is always a compromise between the required yield and quality of the DNA.

Significant efforts to optimize and standardize direct DNA extraction procedures have aimed to improve the reliability of quantitative and qualitative characterization of soil communities, especially bacteria, archaea, fungi and protists (Martin-Laurent et al., 2001; Petric et al., 2011; Plassart et al., 2012; Terrat et al., 2015; Santos et al., 2015; Dimitrov et al., 2017). This includes validation and improvement of a standard method for direct soil DNA extraction (ISO 11063:2012) adopted by the International Organization for Standardization (Phillipot et al., 2010). The method involves three main stages: (1) target cell lysis using the surfactant sodium dodecyl sulphate (SDS) followed by physical disruption by bead beating, (2) protein precipitation by sodium acetate; and (3) nucleic acid precipitation in isopropanol followed by washing in 70% ethanol. This method and its modifications (Plassart et al., 2012; Santos et al., 2015) can be
used for reproducible extraction of DNA from different soil types, which is suitable for soil community analysis using qPCR (Petric et al., 2011) and metabarcoding (Terrat et al., 2015).

Another source of bias in soil community analyses following direct DNA extraction is the effect of co-extraction of substances which inhibit DNA amplification, such as humic acids. ISO 11063 does not cover purification of the extracted DNA. There are several commercial kits available for extraction and purification of DNA from soils and their reliability has been compared (Fredericks et al., 2005; Dineen et al., 2010; Inceoğlu et al., 2010). Extraction kits, which have been most widely used in recent research studies, include the PowerSoil DNA Isolation kit (MO BIO Laboratories, Inc., Carlsbad, CA, USA) and the FastDNA SPIN kit for Soil (MP Biomedicals, LLC, Solon, OH, USA). Additional DNA purification is usually also required for elimination of PCR inhibitors, usually by treatment with polyvinylpolypyrrolidone (PVPP) (Frostegard et al., 1999; Zhou et al., 1995) and/or hexadecyltrimethylammonium bromide (CTAB) (Cho et al., 1996; Malik et al., 1994; Zhou et al., 1995). Braid et al. (2003) also showed that addition of AlNH₄(SO₄)₂ during extraction significantly reduced the co-purification of PCR inhibitors with minimal loss of DNA yield.

Whilst commercial extraction and purification kits are useful for small-scale research, they are usually only suitable for small samples of soil (<10g) and their use in routine soil analyses is further limited by their slow speed of sample throughput and relatively high cost. Ophel-Keller et al. (2008) developed a method capable of extracting DNA from soil samples of 500g with a throughput of 160 samples per day and at a cost of less than 20% of the cost using commercial kits. However, the full details of this method remain unpublished. In the meantime, additional methods suitable for larger scale direct extractions of bacterial and fungal DNA for qPCR analysis of soil-borne fungi and bacteria have been developed (Brierley et al., 2009; Woodhall et al., 2012). Methods suitable for extraction of extracellular eDNA from larger soil volumes have also been described (Taberlet et al., 2012). In this case, a gentler extraction protocol is used to ensure DNA quality is suitable for metabarcoding analysis, which requires higher quality DNA (more purified and fragments longer than 500 bp) than for qPCR analysis. In this case, DNA from 4 kg composite soil samples is extracted by thorough mixing in saturated 0.12M phosphate buffer (pH 8) followed by centrifugation to remove suspended material. DNA is then extracted from aliquots of the supernatant using a commercial kit. Suggested protocols for sampling and extraction of total DNA for qPCR analysis and eDNA for metabarcoding are presented in Appendices 1 and 2.

2.4. Targeted analyses using quantitative polymerase chain reaction (qPCR)

The development of PCR has facilitated major advances in the assessment of the presence of specific target organisms or their functional genes in complex environmental samples, including water, sediments, soils, composts and manures. Real-time quantitative PCR (qPCR) is a safer
technology than conventional PCR since it does not require the use of mutagenic intercalating dyes to stain the DNA. It is also more sensitive, more reproducible, quicker and more cost-effective and is therefore now the most used molecular technique to quantify target sequences. Numerous qPCR approaches have been developed for the detection and enumeration of different organisms in various systems and the subject has been extensively reviewed (Schena et al. 2004; Okubara et al. 2005; Mumford et al. 2006; Cooke et al. 2007; Vincelli and Tisserat 2008; O’Brien et al. 2009; von Felten et al., 2010; Bilodeau 2011; Schena et al., 2013; Sanzani et al., 2014).

The main advantages of qPCR are its sensitivity, ease of use and capacity to run large sample numbers in high-throughput automated testing for quantitative estimation of DNA targets. Its limitations include the variability of the distribution of soil organisms and microorganisms, the efficiency of extraction and purification of DNA from environmental samples and the amount and variation of PCR inhibitors that can be co-extracted from environmental samples. Furthermore, the heterogeneity of PCR templates amplified from different target organisms can create artefacts and biases when attempting to compare populations of different target organisms in the same sample or of the same organism in different samples (Goyer and Dandie, 2012). The accuracy of quantification is also influenced by the number of copies of the target gene in the genome of each organism. Furthermore, since qPCR assays usually target total DNA, there is not always a relationship between DNA quantification and the viability of the detected target. Care is therefore needed when interpreting results in terms of the potential risks or benefits associated with the detection of target organisms or functional genes, which may not be viable or actively expressed. This problem can be tackled by using reverse transcriptase qPCR to detect mRNA targets that are short lived outside of the viable cell, or by using intercalating dyes such as propidium monoazide (PMA) or ethidium monoazide (EMA) which penetrate dead cells and intercalate with the DNA, preventing amplification by PCR (Fittipaldi et al., 2011). The extent to which these methods can be used on DNA extracted from soil and the cost implications for routine analysis have yet to be determined.

A variety of markers have been used to develop specific qPCR assays, which are described in more detail below. Whilst these are often based on unique taxonomic or functional gene markers, sequenced characterised amplified regions (SCAR) have also frequently been used (Gobbin et al., 2007; Hermosa et al., 2001; Holmberg et al., 2009). In these cases, unique DNA bands are generated from genomic DNA of the target organism using DNA fingerprinting techniques such as random amplified polymorphic DNA-PCR (RAPD-PCR), repetitive extragenic palindromic PCR (rep-PCR) or restriction fragment length polymorphism (RFLP). Purification and sequencing of unique DNA fragments then allows soft-ware assisted selection of primers and probes for qPCR assay development and validation. Increasingly, it is now possible to generate specific markers from whole genome comparisons using bioinformatic pipelines which terminate
in the generation of unique primer/probe combinations which are suitable for use in qPCR assays (Pritchard et al., 2013).

2.5. Methods suitable for soil community analyses

Molecular approaches for the analysis of the functions and diversity of whole soil communities have evolved rapidly in the last 10 years and have been widely reviewed and discussed (e.g. Sørensen et al., 2009; Orgiazzi et al., 2015; Drummond et al., 2015; Creer et al., 2016). In addition to qPCR targeting taxon-specific barcode sequences (see above), various other molecular approaches have been used in soil community research, including the following:

2.5.1. Microarrays

The attachment of labelled DNA probes of known sequences to a solid slide or “DNA chip” enables probing with fragments obtained from soil communities. Detection of hybridization between the immobilised DNA probes and the soil DNA, then indicates the array of genes associated with the soil community. For example, development of the microarray GeoChip, with up to 82000 probes covering 141 995 coding sequences from 410 functional gene families, has allowed investigation of the ecological functions of soil microbial communities involved in nitrogen, carbon, sulphur, and phosphorus cycles as well as energy metabolism, antibiotic resistance, metal resistance/reduction, organic remediation, stress responses, bacteriophage, and virulence (He et al., 2010; Tu et al., 2014).

2.5.2. DNA fingerprinting methods

These methods combine PCR amplification of conserved marker genes across a soil community and differentiation between the resulting PCR amplicons according to variation in their DNA sequences. For example, when applied to 16S rRNA gene markers, the method allows the dissection of microbial communities at the level of the phylogeny of their constituents (Smalla et al., 2007). Similar approaches have been used to estimate nematode diversity using the 18s rDNA gene (Waite et al., 2003; Foucher et al., 2004). Two fingerprinting approaches have been most widely used:

- **Density gradient gel electrophoresis (DGGE)** separates the double stranded DNA PCR amplicons during electrophoresis in a polyacrylamide gel with an increasing density gradient of DNA denaturing agents (usually urea and formamide). Fingerprint patterns are formed in the gel when amplicons varying in sequence are denatured at different gradients and their migration is slowed down or stops at different distances during their migration along the gel.

- **Terminal restriction fragment length polymorphism (T-RFLP)** assesses variation in the sequence of PCR amplified DNA by treating with restriction enzymes that cleave the
amplicons into several fragments at specific points in the sequence. The use of fluorescently labelled PCR primers enables the terminal fragments to be separated according to their size in a capillary sequencer.

Chen et al. (2010) compared these methods to other molecular methods, including qPCR and DNA barcoding, for monitoring soil nematodes as biological indicators. They concluded that T-RFLP was well suited for routine monitoring purposes based on the advantages of high-throughput, ease of comparison between samples and rapid data analysis. T-RFLP has also been used to survey the bacterial biogeography of British soils (Griffiths et al., 2011). However, it has been shown that such fingerprinting techniques sometimes only consider the most abundant phylotypes and so diversity estimates can be poorly correlated with true community diversity (Lalande et al., 2013).

2.5.3. Next generation high throughput sequencing

While PCR-based methods have been widely used to study community interactions at high taxonomic levels, the specificity of available primers and the labour and expense involved in sequencing or otherwise differentiating PCR amplicons has limited both sample throughput and the resolution to which individuals can be identified within the soil community. With the development of the first true high throughput sequencing (HTS) 454 Life Sciences platform (Margulies et al. 2005), it became possible to pyrosequence millions of individual amplified molecules in parallel. Subsequent developments in sequencing technology have further increased the depth of sequencing and opportunities for high sample throughput (Loman et al. 2012). Three new next generation sequencing (NGS) platforms were released in 2011 alone: Ion Torrent’s PGM, Pacific Bioscience’s RS and the Illumina MiSeq (Quail et al., 2012). Alongside these advances in technology are developments of bio-informatics tools that manage large data flows, compare data with specialised databases and extract relevant information, creating new perspectives for investigating the soil microbiome (Uroz et al., 2013). Illumina sequencing-bysynthesis has particularly enabled greater sequencing depth and higher sample throughput alongside reduced costs. More recently, single molecule sequencing technologies, such as Pacific Biosystems and Oxford Nanopore, have allowed the generation of much longer reads from samples where DNA is only present at low concentrations. However, higher costs, reduced throughput and increased error rates of the latest platforms mean that Illumina currently remains the platform of choice for community ecology research (Schmidt et al., 2013; Creer et al., 2016).

High throughput sequencing technology is revolutionising the way in which the functions and diversities of soil communities are investigated through analysis of directly or indirectly extracted total DNA or of the eDNA. Three main approaches are currently being followed:
• **Metabarcoding** involves PCR amplification of pre-selected barcode sequences in phylogenetically relevant marker genes followed by simultaneous high throughput sequencing of all amplicons generated and bioinformatic sequence comparisons using databases of verified specimen sequences. Several sequence databases have been employed, including RDP (Cole *et al*., 2014), Greengenes (DeSantis *et al*., 2006), SILVA (Pruesse *et al*., 2007), UNITE (Abarenkov *et al*., 2010), BOLD (Ratnasingham & Hebert, 2007), & Genbank (Benson *et al*., 2012). First used to describe meiofauna community structures by analysing eDNA in sediments (Creer *et al*., 2010), metabarcoding methods were also developed to analyse soil for DNA from fungi, bryophytes, enchytraeids, beetles and even birds (Epp *et al*., 2012). High-throughput metabarcoding studies on fungi and other eukaryotic microorganisms are rapidly becoming more frequent and more complex, and several new bioinformatic pipelines have been described for metabarcoding bacteria (Gellie *et al*., 2017), fungi (Bálint *et al*., 2014) and fauna (Yang *et al*., 2013; de Groot *et al*., 2016) including protists (Geisen *et al*., 2015), nematodes (Posazinska *et al*., 2010 and 2010b; Sapkota and Nicolaisen, 2015) and earthworms (Bienert *et al*., 2012).

• **Metagenomics** entails random (shotgun) sequencing of long strands of soil DNA to elucidate the taxonomic structure and potential functional genomic capability of a community. In contrast to metabarcoding, metagenomics does not require an initial PCR step, thus avoiding potential biases associated with the use of different primer sets with varying amplification efficiencies (Logares *et al*., 2013). Shotgun sequencing provides an amplification independent method for assessing community diversity, additionally allowing for the capture of information from groups that are otherwise difficult to survey (Narasingarao *et al*., 2012). For example, Delmont *et al*. (2012) described the metagenome of a Rothamsted grassland soil using 454 pyrosequencing and showed that only 1% of the annotated sequences found corresponded to known sequenced genomes. Unexpectedly low seasonal and vertical soil metagenomic functional class variations were also observed. Metagenomic analysis of soils across global ecosystems is indicating major differences in soil microbiomes in terms of both taxonomic and functional representation (Noronha *et al*., 2017). Differences in functional and/or taxonomic diversity are also shown to vary with agricultural land use (Manoharan *et al*., 2017), including tillage and crop management practices (Souza *et al*., 2015), pH and fertilizer adjustment (Fierer *et al*., 2012; Zhalnina *et al*., 2015) and the use of organic amendments (Jenkins *et al*., 2017). Significant investment is being targeted towards sequence-based approaches to understand the soil microbiome. For example, The Earth Microbiome Project ([www.earthmicrobiome.org](http://www.earthmicrobiome.org)) alone is characterizing 200 000 samples from researchers all over the world. Results of metagenomic analysis are known to vary with the DNA extraction method, database choice and the annotation procedure (Jacquiod *et al*., 2017).
In support of this, improved curated genomic databases, such as RefSoil (Choi et al., 2016) and FuSiON (Zhu et al., 2015), are being developed to provide a soil-specific framework with which to annotate and understand soil sequencing projects.

- **Metatranscriptomics** allows study of the actual functional activity of the entire soil microbiome through shotgun sequencing of the messenger RNA (mRNA) (Carvalhais and Schenk, 2013; Myrold et al., 2014; Thies, 2015). Presence of mRNA reflects the portion of the soil microbial community that is active at the time of sampling (Blagodatskaya and Kuzyakov, 2013) and thus provides a robust means to examine biological responses to soil management. The process is challenging as mRNA typically represents less than 5% of the RNA extracted (Carvalhais et al., 2013) and is often extremely short-lived. The mRNA in the extract must be enriched and the rRNA removed. Subtractive hybridization, treatment with endonucleases that preferentially degrade rRNA or duplex specific nuclease treatment are used to remove rRNA (Yi et al., 2011). Eukaryotic mRNA can be separated by binding the 3’-poly-A tails of the transcripts to surfaces coated with poly(dt)probes, thus enriching mRNA from bacteria and archaea in the extract. The mRNA is then reverse-transcribed (RT) into cDNA for high-throughput sequencing.

### 2.6. Molecular markers for bio-indicators of soil health

The size and diversity of the biological communities sustained within different soils can be determined according to the prevalence of molecular markers within their DNA, which have been selected to identify different taxonomic groups, functional groups or individuals within the community. Ritz et al. (2009) listed several bio-indicators that can be targeted in this way when assessing the overall health status of soils in relation to crop productivity and long-term sustainable land management:

#### 2.6.1. Biological communities contributing to soil biodiversity

For taxonomic barcode markers within bacteria and archaea, highly conserved target sequences within the 16S ribosomal RNA gene have been most widely used (Fierer et al., 2005; Sogin et al., 2006), although hypervariable regions within the gene and the 16-23S intergenic spacer have also been used for more taxa-specific analyses (e.g. Becker et al., 2000; Suzuki et al., 2000; Takai & Horikoshi, 2000 and Blackwood et al., 2005). Various housekeeping genes have also proven useful as taxonomic barcodes since they are essential and are therefore not lost from genome, but evolve more quickly than 16S rDNA. The most commonly used taxonomic markers used to identify soil-borne fungi and oomycetes also often include rDNA markers, including the intergenic transcribed regions ITS1 and ITS2, located between the small sub unit (SSU) 18S and the large sub unit (LSU) 28S genes and separated by the 5.8S gene (Schoch et al., 2012; Schena et al., 2013). For other taxonomic
groups, a diverse set of loci from the analogous eukaryotic rRNA gene array (e.g. ITS, 18S or 28S rRNA) (Bik et al., 2012a; McGuire et al., 2010; Epp et al. 2012) have been employed. Barcoding of nematodes has mainly relied on rDNA targets, including ITS1, ITS2, 18S, D2-D3 expansion segments of the 28S rRNA gene, 5S and intergenic spacer (IGS) targets, although other targets also include the mitochondrial cytochrome oxidase gene (CO1) (Blok, 2010). Some 51 taxon-specific markers based mainly on the rDNA locus (Floyd et al., 2002; Griffiths et al., 2006; Holtermann et al., 2006; Rybarczyk-Mydlowska et al., 2012; Vervoort et al., 2012) have been used for characterisation of nematode assemblages (Quist et al. 2016 and 2017). Similarly, for protists, conserved sequences within the 18S, ITS and CO1 genes are often used (Pawlowski et al. 2012). The 18S and CO1 genes are also widely used for barcoding meiofauna and macrofauna (Hebert et al., 2003; Deagle et al., 2014) together with additional sequence information from the 12S and 16S genes for macrofauna (Epp et al., 2011). Capra et al. (2016) also recently described a new set of markers based on 18S rRNA for metabarcoding of soil metazoa. In general, markers based on rDNA often provide useful sequence barcodes with good resolution at genus or higher, although resolution to species level is not always possible. Furthermore, quantification based on total gene copies detected can be variable due to the different numbers of rRNA copies associated with different species (Klappenbach et al., 2000).

2.6.2. Microbial activity affecting soil nutrient availability/retention

Organisms involved with biogeochemical processes such as carbon, nitrogen, phosphate and sulphur cycling are important indicators of soil health. Genes controlling these processes have been used as markers to quantify functional groups of these organisms in soils. Recent examples describing the use of qPCR to monitor presence and quantify various functional microbial groups according to their expected activities in soils and sediments are shown in Table 1.

For carbon cycling organisms, these include genes expressing enzymes involved in oxidation of organic matter to CO₂ (soil respiration), which include β-glucosidases, cellulases and phenol oxidases, as well as the consumption of methane by methanotrophic bacteria (methane oxidases). Functional groups of bacteria can also be quantified using marker genes for enzymes involved in nitrogen cycling, including:

- **Nitrogenases** in nitrogen-fixing *Rhizobium* species and cyanobacteria (blue green algae), which assimilate atmospheric nitrogen into ammonia.
- **Ureas and amydases** in nitrogen mineralising bacteria, which convert organic nitrogen to ammonium ions.
- **Ammonium oxidases** in nitrifying bacteria, which produce nitrite and nitrate from ammonium.
- **Nitrate reductases**, nitrite reductases and nitrous oxide reductases in denitrifying bacteria, which release gaseous nitrous oxide and nitrogen from nitrates and nitrites.

Similarly, functional groups of bacteria and fungi contributing to phosphate and sulphur cycles have been quantified according to marker genes expressing the key enzymes involved. Acid and alkaline phosphatase enzymes produced by bacteria, fungi and plant roots serve to transform complex and sometime unavailable forms of organic P into assimilable phosphate. Population dynamics of sulphate-reducing bacteria have similarly been quantified by quantification of their functional genes responsible for dissimilatory reduction of sulphates to adenosine 5’-phosphosulfate (APS), sulphites and eventually to gaseous hydrogen sulfide under anaerobic conditions.
Table 1: Marker genes used to monitor microbial activity affecting soil nutrient cycling

<table>
<thead>
<tr>
<th>Biological function</th>
<th>Targeted genes</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>C-cycling</td>
<td>Ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO) (\textit{cbbL})</td>
<td>Powell \textit{et al.}, 2006</td>
</tr>
<tr>
<td></td>
<td>Methane mono-oxygenase (\textit{pmoA})</td>
<td>Selesi \textit{et al.}, 2007</td>
</tr>
<tr>
<td></td>
<td>Methyl coenzyme M reductase (\textit{mcrA})</td>
<td>Yergaeu \textit{et al.}, 2009</td>
</tr>
<tr>
<td></td>
<td>Cellulase (\textit{cel})</td>
<td>Freitag \textit{et al.}, 2010</td>
</tr>
<tr>
<td></td>
<td>Chitinase (\textit{chiA})</td>
<td>Pereyra \textit{et al.}, 2010</td>
</tr>
<tr>
<td></td>
<td>Alkane mono-oxygenase (\textit{alkb})</td>
<td>Yergeau \textit{et al.}, 2012</td>
</tr>
<tr>
<td></td>
<td>PAH ring hydroxylating dioxygenase (PAH-RHD)</td>
<td></td>
</tr>
<tr>
<td>N-cycling</td>
<td>Ammonia mono-oxygenase (\textit{amoA})</td>
<td>Hai \textit{et al.}, 2009</td>
</tr>
<tr>
<td></td>
<td>Nitrogenase reductase (\textit{nifH})</td>
<td>Hayden \textit{et al.}, 2010</td>
</tr>
<tr>
<td></td>
<td>Nitrous oxide reductase (\textit{nosZ})</td>
<td>Bru \textit{et al.}, 2011</td>
</tr>
<tr>
<td></td>
<td>Nitrite reductase (\textit{nirS}, and \textit{nirK})</td>
<td>Dose \textit{et al.}, 2015,</td>
</tr>
<tr>
<td></td>
<td>Nitrate reductase (\textit{narG} and \textit{napA})</td>
<td>Dandie \textit{et al.}, 2011</td>
</tr>
<tr>
<td></td>
<td>Nitric oxide reductase (\textit{norB})</td>
<td></td>
</tr>
<tr>
<td>S-cycling</td>
<td>Dissimilatory sulfite reductase (\textit{dsrA})</td>
<td>Ben-Dov \textit{et al.}, 2007</td>
</tr>
<tr>
<td></td>
<td>Adenosine 5'-phosphosulfate reductase (\textit{aprA}, \textit{ApsA})</td>
<td>Blazejak &amp; Schippers, 2011</td>
</tr>
<tr>
<td>P-cycling</td>
<td>acid phosphatase (\textit{phoC})</td>
<td>Fraser \textit{et al.}, 2017</td>
</tr>
<tr>
<td></td>
<td>alkaline (\textit{phoD}) phosphatase</td>
<td></td>
</tr>
</tbody>
</table>
2.6.3. Microbial symbionts contributing to crop nutrition

**Arbuscular mycorrhizal fungi**
Primers and probes to quantify abundance of different arbuscular mycorrhizal fungal (AMF) taxa in roots and in soil have been selected from taxon-specific markers in the nuclear large ribosomal subunit RNA genes (nrDNA) and the mitochondrial ribosomal mtDNA (König et al., 2010; Thonar et al., 2012; Voříšková et al., 2017). Interpretation of the qPCR results has been complicated by the multinuclear and multigenomic cellular organization of these fungi and the high DNA sequence diversity within the smallest biologically relevant units (i.e. single-spore isolates). Markers based on nrDNA, rather than mtDNA, are thought to be more suitable for the quantification of multiple AMF taxa as copy numbers of the former are better related to fungal biomass across taxa.

**Nitrogen-fixing bacteria**
Molecular markers used to quantify soil populations of nitrogen-fixing bacteria, such as *Rhizobium, Bradyrhizobium* and *Sinorhizobium* spp., include *rpoE1, nodC, nodD* and *nodZ* genes (Trabelsi et al., 2009; Boonen et al., 2010; Furseth et al., 2010; Macdonald et al., 2011). Nitrogen-fixing bacteria belonging to the genus *Frankia* have also been studied using markers within target genes *nifH* (Samant et al., 2012) and 23S rRNA (Samant et al., 2014; Ben Tekaya et al., 2017). Strain specific and general primer/probe sets are available for molecular qPCR analysis targeting these genes, offering a rapid and comparable alternative to the laborious procedure of most-probable number bioassays based on counting the number of nodules on the roots of test plants. Nodulation potential can be overestimated by the molecular methods due to presence of dead cells and DNA, which can remain in soil for up to 5 months after inoculation.

**Plant growth promoting rhizobacteria (PGPR)**
Several molecular markers have been used to detect and quantify bacteria considered to have some activity, which is plant growth promoting in the rhizosphere of different crops (Table 2). In most cases, strain-specific assays have been used to determine the fate of PGPR strains added to soils during assessment of their suitability for commercialisation. Functional genes of fluorescent *Pseudomonas* spp. have been used as markers for grouping strains with antibiotic or hydrogen cyanide activities which are dispersed across several different species (Kim et al., 2013).
Table 2: Marker genes used to monitor plant growth promoting bacteria (PGPR) in soils

<table>
<thead>
<tr>
<th>PGPR</th>
<th>Marker</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Azospirillum lipoferum</td>
<td>CRT1 (RFLP product SCAR marker)</td>
<td>Couillerot et al., 2010</td>
</tr>
<tr>
<td>Azospirillum brasilense</td>
<td><em>nifA</em> (nitrogen fixation)</td>
<td>Faleiro et al., 2013</td>
</tr>
<tr>
<td>Bacillus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Herbaspirillum seropedicae</td>
<td>HERBASI1 (prophage sequence)</td>
<td>Pereira et al., 2014</td>
</tr>
<tr>
<td>Enterobacter radicincitans</td>
<td>16S rRNA</td>
<td>Schreiner et al., 2009</td>
</tr>
<tr>
<td>Paenibacillus polymyxa</td>
<td>16S rRNA</td>
<td>Timmusk et al., 2009</td>
</tr>
<tr>
<td>Fluorescent Pseudomonas spp.</td>
<td>Multiple strain specific markers</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>(RAPD product SCAR markers)</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>phlD</em> (2,4-diacetylphloroglucinol)</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>phzCD</em> (phenazine-1-carboxylic acid)</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>hcnBC</em> (hydrogen cyanide)</td>
<td></td>
</tr>
<tr>
<td>Pseudomonas brassicacearum</td>
<td>OPA2-73 (RAPD product SCAR marker)</td>
<td>Holmberg et al., 2009</td>
</tr>
<tr>
<td>Pseudomonas sp. (DSMZ 13134)</td>
<td><em>dnaX</em> (housekeeping gene)</td>
<td>Mosiman et al., 2016</td>
</tr>
</tbody>
</table>

2.6.4. Biocontrol agents

Molecular markers have been used for detection and quantification of commonly used fungal biocontrol agents to assess their establishment, distribution and longevity in different soils. General ITS markers have often been used for fungi such as the mycoparasitic *Trichoderma harzianum* (Lopez-Mondéjar et al., 2010). RNA detection using reverse transcriptase amplification of the same markers has enabled estimation of viable populations (Beaulieu et al., 2011). Strain specific markers within the housekeeping gene *aox1* allow differentiation of commercial biocontrol agent strains of *T. harzianum* from other soil inhabiting strains (Horn et al., 2016). Other markers have been used for analysis of *T. atroviride* (Cordier et al., 2007; Savazzini et al., 2008). Vallance et al. (2009) used ITS markers to study the biocontrol agent *Pythium oligandrum* in the rhizosphere. ITS markers have also been used to study the nematophagous fungus *Plectosphaerella cucumerina* used for control of potato cyst nematodes (Atkins et al., 2003) and the entomopathogenic fungi *Entomophaga maimaiga* (Castrillo et al., 2007) and *Beauveria bassiana* (Bell et al., 2009; Garrido-Jurado et al., 2016). SSR (single sequence repeat) microsatellite markers have also been used to distinguish between biocontrol strains of *Beauveria bassiana* and *B. brongniartii* (Canfora et al., 2016). SCAR (sequence-characterised amplified region) markers, based on unique RAPD amplified sequences, have also been used to recognise the biocontrol agents *B. bassiana* (Castrillo et al., 2008) and *Chaetomium globosum* (Aggarwal et al., 2014).
Detection and identification of soilborne bacterial biocontrol agents, such as those belonging to the *Bacillus subtilis* group (including *Bacillus subtilis* and *B. amyloliquefaciens*), have also been based on SCAR markers, identified from unique PCR amplicons generated after RAPD or rep-PCR analysis. Strain specific differentiation has been based on sequence variation within housekeeping genes, including a tryptophan biosynthesis gene *trpE* (Johansson *et al.*, 2014) and RBAM 007760 (Gotor-Vila *et al.*, 2016), a gene involved in surface adhesion and biofilm formation. This has allowed identification of strains, which better colonize the rhizosphere. Strains of *Pseudomonas fluorescens* that produce the antibiotic 2,4-diacetylphloroglucinol (2,4-DAPG) have been monitored in cereal rhizospheres using *phlD* gene sequences as markers (Mavrodi *et al.*, 2007).

### 2.6.5. Nematode assemblages

For community analysis of nematodes extracted from soils, markers within the small sub unit (SSU) of the ribosomal DNA (rDNA) have been most widely exploited to identify nematode feeding guilds (Griffiths *et al.*, 2006). Donn *et al.*, (2012) developed a directed terminal restriction fragment length polymorphism (dT-RFLP) method whereby PCR primers are used to amplify the SSU rDNA, followed by restriction of the amplicons with selected enzymes and comparison of the resulting fragment sizes with those in sequence databases produced from fully characterised species. A series of rDNA barcodes have been selected that allow identification to family and genus levels (Floyd *et al.*, 2002; Holterman *et al.*, 2006; Vervoort *et al.*, 2012), which can then be allocated to feeding guilds. These have been used experimentally to show effects of long term soil management practices on the frequency of occurrence of each taxon following qPCR analyses with up to 51 primer sets (Quist *et al.*, 2016 and 2017). Wang (2012) proposed a simpler system based on qPCR of 18S rDNA targets, but which only identifies certain guilds of bacterial-feeding, fungal-feeding, omnivorous and predatory nematodes but not herbivorous nematodes. A range of bespoke markers have also been designed for specific detection of different plant pathogenic nematode species (e.g. Madani *et al.*, 2005; Holeva *et al.*, 2006; Nakhla *et al.*, 2010).

### 2.6.6. Soil-borne plant pathogens

Bilodeau (2011) reviewed the use of qPCR methods for detection of soil-borne plant pathogenic micro-organisms and nematodes. An updated list of available qPCR assays suitable for detection of soil-borne pathogens is shown in Table 3.
Table 3: Markers used for detection and quantification of soil-borne pathogens

<table>
<thead>
<tr>
<th>Pathogen</th>
<th>Marker</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aphanomyces cochlioides</td>
<td>rDNA (ITS)</td>
<td>Almquist et al., 2016</td>
</tr>
<tr>
<td>Colletotrichum coccodes</td>
<td>rDNA (ITS)</td>
<td>Cullen et al., 2002</td>
</tr>
<tr>
<td>Fusarium graminearum</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fusarium culmorum</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fusarium avenaceum</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fusarium poae</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fusarium culmorum</td>
<td>PKS13 Zearalenone biosynthesis</td>
<td>Atoui et al., 2012</td>
</tr>
<tr>
<td>Fusarium oxysporum</td>
<td>rDNA (ITS)</td>
<td>Cullen et al., 2005</td>
</tr>
<tr>
<td>Fusarium oxysporum f. sp. melonis</td>
<td>translation elongation factor (TEF-1α)</td>
<td>Haegi et al., 2013</td>
</tr>
<tr>
<td>Fusarium culmorum</td>
<td>Trichodiene synthase gene (tri5)</td>
<td>Hogg et al., 2010</td>
</tr>
<tr>
<td>Globodera rostochiensis</td>
<td>rDNA (ITS1)</td>
<td>Filion et al., 2003</td>
</tr>
<tr>
<td>G. pallida</td>
<td>rDNA (ITS1)</td>
<td>Toyota et al., 2008</td>
</tr>
<tr>
<td>Gaeumannomyces graminis var.</td>
<td>rDNA (TS1)</td>
<td>Bithell et al., 2012</td>
</tr>
<tr>
<td>Avenae</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gaeumannomyces graminis var.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tritici</td>
<td>translation elongation factor gene (EF1-α)</td>
<td>Keenan et al., 2015</td>
</tr>
<tr>
<td>Helminthosporium solani</td>
<td>rDNA (ITS)</td>
<td>Cullen et al., 2001</td>
</tr>
<tr>
<td>Heterodera avenae</td>
<td>mitochondrial cytochrome oxidase subunit 1 (COI)</td>
<td>Touni et al., 2015</td>
</tr>
<tr>
<td>Heterodera schachtii</td>
<td>rDNA (ITS)</td>
<td>Madani et al., 2005</td>
</tr>
<tr>
<td>Phomopsis sclerotiorides</td>
<td>rDNA (ITS)</td>
<td>Shishido et al., 2013</td>
</tr>
<tr>
<td>Phytophthora kernoviae</td>
<td>rDNA (ITS)</td>
<td>Hughes et al., 2011</td>
</tr>
<tr>
<td>Phytophthora rubi</td>
<td>cox1 cytochrome oxidase subunit 1</td>
<td>Woodhall &amp; Peters, 2014</td>
</tr>
<tr>
<td>Plasmodiophora brassicae</td>
<td>rDNA (18S and ITS1)</td>
<td>Sundelin et al., 2010</td>
</tr>
<tr>
<td></td>
<td>rDNA (18S)</td>
<td>Rennie et al., 2011</td>
</tr>
<tr>
<td></td>
<td>rDNA (ITS)</td>
<td>Wallenhamer et al., 2012</td>
</tr>
<tr>
<td></td>
<td>rDNA (ITS)</td>
<td>Li et al., 2013</td>
</tr>
<tr>
<td></td>
<td>rDNA (ITS)</td>
<td>Kennedy et al., 2013</td>
</tr>
<tr>
<td></td>
<td>rDNA (18S)</td>
<td>Cao et al., 2014</td>
</tr>
<tr>
<td></td>
<td>rDNA (ITS1)</td>
<td>Deora et al., 2015</td>
</tr>
<tr>
<td>Polymyxa betae</td>
<td>rDNA (ITS)</td>
<td>Ward et al., 2004</td>
</tr>
<tr>
<td>Polymyxa graminis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pathogen</td>
<td>Marker</td>
<td>Reference</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>---------------------------------------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td><em>Polyscytalum pustulans</em></td>
<td>rDNA (ITS)</td>
<td>Lees et al., 2009</td>
</tr>
<tr>
<td><em>Pratylenchus penetrans</em></td>
<td>β-1,4-endoglucanase gene</td>
<td>Mokrini et al., 2013</td>
</tr>
<tr>
<td><em>Pythium violae</em></td>
<td>rDNA (ITS)</td>
<td>Cullen et al., 2007</td>
</tr>
<tr>
<td></td>
<td>rDNA (ITS and 5.8S)</td>
<td>Schroeder et al., 2006</td>
</tr>
<tr>
<td><em>Rhizoctonia cerealis</em></td>
<td>β-tubulin gene</td>
<td>Guo et al., 2012</td>
</tr>
<tr>
<td></td>
<td>Unique SCAR sequence</td>
<td>Woodhall et al., 2017</td>
</tr>
<tr>
<td><em>Rhizoctonia solani</em> AG1-1</td>
<td>rDNA (ITS), β-tubulin</td>
<td>Lees et al., 2002</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Budge et al., 2009</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Woodhall et al., 2013</td>
</tr>
<tr>
<td><em>Rhizoctonia solani</em> AG2-1</td>
<td>rDNA (ITS), β-tubulin</td>
<td>Abbasi et al., 2014</td>
</tr>
<tr>
<td><em>Rhizoctonia solani</em> AG2-2 IIIB</td>
<td>rDNA (ITS1, 18S and 5.8S)</td>
<td>Rogers et al., 2009</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kim &amp; Knudsen, 2008</td>
</tr>
<tr>
<td><em>Sclerotinia sclerotiorm</em></td>
<td>mitochondrial small subunit</td>
<td>Rogers et al., 2009</td>
</tr>
<tr>
<td></td>
<td>rRNA intron and ORF1</td>
<td>Qu et al., 2011</td>
</tr>
<tr>
<td><em>Sclerotium cepivorum</em></td>
<td>rDNA (ITS)</td>
<td>Woodhall et al., 2012</td>
</tr>
<tr>
<td><em>Spongospora subterranea</em></td>
<td>rDNA (ITS)</td>
<td>van de Graaf et al., 2003</td>
</tr>
<tr>
<td></td>
<td>rDNA (ITS2)</td>
<td>Qu et al., 2011</td>
</tr>
<tr>
<td><em>Streptomyces spp.</em></td>
<td>nec1 virulence gene</td>
<td>Qu et al., 2011</td>
</tr>
<tr>
<td></td>
<td>txtAB thaxtomin toxin synthetase gene</td>
<td>Cullen &amp; Lees, 2007</td>
</tr>
<tr>
<td></td>
<td>16S rRNA</td>
<td>Schlater et al., 2010</td>
</tr>
<tr>
<td><em>Synchytrium endobioticum</em></td>
<td>rDNA (ITS)</td>
<td>Van Gent-Pelzer et al., 2010</td>
</tr>
<tr>
<td><em>Verticillium dahliae</em></td>
<td>β-tubulin gene</td>
<td>Debode et al., 2011</td>
</tr>
<tr>
<td><em>Verticillium longisporum</em></td>
<td>rDNA (ITS)</td>
<td>Bilodeau et al., 2012</td>
</tr>
<tr>
<td><em>Verticillium tricorpus</em></td>
<td></td>
<td>Peters, 2012</td>
</tr>
<tr>
<td><em>Verticillium albo-atrum</em></td>
<td>rDNA (IGS)</td>
<td>Maurer et al., 2013</td>
</tr>
<tr>
<td><em>Verticillium dahliae</em></td>
<td>β-tubulin gene</td>
<td>Duressa et al., 2012</td>
</tr>
<tr>
<td></td>
<td>rDNA (IGS)</td>
<td>Bilodeau et al., 2012</td>
</tr>
<tr>
<td><em>Verticillium dahliae</em></td>
<td>rDNA (18S)</td>
<td>Banno et al., 2011</td>
</tr>
<tr>
<td><em>Verticillium longisporum</em></td>
<td>rDNA (ITS-5.8S)</td>
<td></td>
</tr>
</tbody>
</table>
3. Conclusions

The analysis of biological community structure and function in soil is benefiting from the availability of an ever-expanding assortment of molecular tools. Although complex and expensive, and therefore mostly confined to use in research, some molecular procedures are rapidly evolving and may become available as affordable procedures for routine analysis of biological indicators of soil health. The use of qPCR to detect and quantify specific organisms from total soil DNA and eDNA is already starting to be used on a routine basis. For example, the Predicta® soil testing service offered by the South Australian Research and Development Institute (SARDI) is already offering quantitative analysis of the distribution of some soil-borne pathogens of broadacre and potato crops in relation to the risk of disease development (Ophel-Keller et al., 2008). The first uses of high throughput next generation sequencing in national and international surveys are suggesting that there is a core microbiome in geographically distant and disparate soils (Orgiazzi et al., 2013), but that there are also major differences in taxonomic and functional representation (Noronha et al., 2017). Metagenomic analysis of soils is also showing that the microbiome is dynamic, varying with season (Jumpponen et al., 2010) and agricultural land use (Fierer et al., 2012; Souza et al., 2015; Zhalnina et al., 2015; Jenkins et al., 2017; Manoharan et al., 2017).

A number of technical challenges remain to be fully overcome in the application of molecular analyses for reliable characterisation of the taxonomy and function of soil communities (Delmont et al., 2012). It is very difficult to suggest standardised practices to sample and extract unbiased and representative samples of DNA from organisms with very different cell membranes and accessible DNA and across the full variability of soil types. This problem is exacerbated by the uneven spatial distribution of microbial communities in soil. Furthermore, the yield and quality of extracted DNA varies with the chemical and physical properties of different soils (Feinstein et al., 2009) as well as the extraction method used (Petric et al., 2011). Although attempts have been made to standardize methods for DNA extraction from soil (Phillipot et al., 2010), their validation has been restricted to certain analytical methods. Since different analytical methods have different DNA quality requirements, there remains a need for further validation and standardisation of methods suitable for use with each new analytical method of choice. In addition to sampling and extraction, biases in soil community analyses can be attributed to differences in the molecular markers targeted as well as in the methods used for their analysis and the databases used for their identification. Heterogeneity of PCR templates amplified from different target organisms can create artefacts and biases (Goyer and Dandie, 2012). The accuracy of quantification is also influenced by the number of copies of the target gene in the genome of each organism.

Whilst standardization of sampling, extraction and analytical methods is highly desirable, it is unlikely that a single standardised procedure will be suitable for all types of molecular soil analysis. It is much more likely that a series of standardised procedures will be needed that are optimised for several parameters, including the types of organism to be studied, the numbers of soil samples to
be tested, the amounts of soil in each sample and the type of analysis to be performed. Of key importance will be the introduction of suitable internal controls that ensure efficient extraction of high quality DNA and the accuracy of its detection, quantification and representation within the soil community. For routine comparisons of soil health, the costs of sampling, DNA extraction and molecular analysis are also likely to be highly influential. High-throughput PCR and sequencing methods contribute to a significant increase in testing efficiencies, allowing simultaneous investigation of multiple targets and whole communities from a single DNA extract. Developments in technology are also moving towards the performance of molecular analyses *in situ*, rather than having to transport samples for laboratory analysis. Nevertheless, the cost of molecular analysis remains a key constraint to its routine application in monitoring soil health.

The wide range of taxonomic and functional markers that are available for probing soil DNA, together with the high throughput methods that are available for automated analysis, mean that multiple analyses can now be easily and simultaneously performed on a single soil DNA extract. Since DNA extraction is the most expensive part of the analysis, it is cost effective to test each extract for multiple markers. The challenge over the current Research Partnership is to identify the most appropriate biological soil health indicators to include in this type of analysis. Knowledge exchange workshops are already in progress, together with growers and agronomists, to explore the most useful targets for routine testing. The challenge will be to link the results of such testing to yield and quality benefits across entire cropping systems. Future workshops will also include demonstrations of the benefits of testing and will involve agronomists in the sampling procedures and, eventually, in interpretation of the results of DNA analyses. Future research within the Partnership will concentrate on validating and standardising the most appropriate molecular methods and building data on the effects of long-term soil management practices on key soil health indicators, including both beneficial and pathogenic organisms.
4. References


43. Couillerot, O., Bouffaud, M-L., Baudoin, E., Muller, D., Caballero-Mellado, J. and Moënne-Loccoz, Y. 2010. Development of a real-time PCR method to quantify the PGPR strain Azospirillum lipoferum CRT1 on maize seedlings. Soil Biology and Biochemistry 42, 2298–2305.


5. Appendix 1

Method for total DNA extraction from soil samples of up to 500g
(after Woodhall et al., 2012)

Scope: This method is suitable for direct extraction and purification of total DNA from composite soil samples of up to 500g. The procedure below is described for samples of 250g.

Materials:
1. Minimix auto paint shaker (Merris Engineering Ltd., Ireland)
3. Grinding buffer (120 mM sodium phosphate buffer pH 8.2, 2% centrimonium bromide, 1.5M sodium chloride)
4. Antifoam B (Sigma-Aldrich)
5. 5M potassium acetate
6. Isopropanol
7. Silicon dioxide (Sigma-Aldrich)
8. Wizard® Magnetic DNA Purification System for Food (Promega)

Procedure:
1. Sieve soil to remove stones.
2. Place 250 g soil into 1000 ml Nalgene wide mouth environmental bottles with 20 stainless steel ball bearings (25.4 mm), 500 ml grinding buffer and 15 ml Antifoam B.
3. Shake for 4 min. in a minimix auto paint shaker (Merris Engineering Ltd., Ireland).
4. Centrifuge a 50 ml sub-sample at 5,000g for 5 min.
5. Transfer 20 ml of the supernatant to a clean tube containing 2 ml of 5M potassium acetate and incubate on ice for 10 min.
6. Centrifuge at 12,000 g for 5 min.
7. Transfer the supernatant to a clean tube containing 15 ml isopropanol and 800 µl silicon dioxide suspension.
8. Shake on flat bed shaker at 100 rpm for 15 min.
9. Centrifuge at 12,000 g for 5 min.
10. Discard supernatant and add 2ml Buffer A (Promega Wizard® Food Kit) to the pelleted silica particles.
11. Shake tubes on their side for 10 min, at 65 °C and 100 rpm.
12. Centrifuge at 12,000 g for 5 min.
13. Extract DNA from 1000 µl of the supernatant by magnetic capture using the Wizard® Magnetic DNA Purification System for Food (Promega) in a Kingfisher ML magnetic particle processor (Thermo Fisher Scientific) according to the manufacturers’ instructions.
6. Appendix 2

Method for extraction of extracellular eDNA from soil samples of up to 500 g
(after Taberlet et al., 2012)

Scope: This method is suitable for direct extraction and purification of eDNA from composite soil samples of up to 500 g. The procedure below is described for samples of 250 g.

Materials:
1. Sodium phosphate buffer pH 8 (1.97 g NaH$_2$PO$_4$ and 14.7 g Na$_2$HPO$_4$ per l sterile distilled H$_2$O).
2. NucleoSpin Soil kit (Macherey-Nagel, Düren, Germany)

Procedure:
1. Sieve soil to remove stones
2. Add an equal weight of phosphate buffer (pH 8) to the soil in a wide neck bottle.
3. Mix thoroughly by gentle orbital shaking for 15-30 min (e.g. at 50-100 rpm) to homogenise the sample.
4. Centrifuge aliquots of the resulting soil suspension at 10,000 rcf for 10 min.
5. Purify DNA from 500 µl of the supernatants using a Nucleospin Soil commercial kit (Macherey-Nagel, Düren, Germany) following the manufacturers’ instructions but skipping their lysis step.