Impact of metagenomic DNA extraction procedures on the identifiable endophytic bacterial diversity in Sorghum bicolor (L. Moench)

Mapula Kgomotso Annah Maropola a, Jean-Baptiste Ramond a,b, Marla Trindade a,*

a Institute for Microbial Biotechnology and Metagenomics (IMBM), Biotechnology Department, University of the Western Cape, Bellville 7535, Cape Town, South Africa
b Center for Microbial Ecology and Genomics (CMEG), Genomics Research Institute, University of Pretoria, South Africa

Abstract

Culture-independent studies rely on the quantity and quality of the extracted environmental metagenomic DNA (mDNA). To fully access the plant tissue microbiome, the extracted plant mDNA should allow optimal PCR applications and the genetic content must be representative of the total microbial diversity. In this study, we evaluated the endophytic bacterial diversity retrieved using different DNA extraction procedures. Metagenomic DNA from sorghum (Sorghum bicolor L. Moench) stem and root tissues were extracted using two classical DNA extraction protocols (CTAB- and SDS-based) and five commercial kits. The mDNA yields and quality as well as the reproducibility were compared. 16S rRNA gene terminal restriction fragment length polymorphism (t-RFLP) was used to assess the impact on endophytic bacterial community structures observed. Generally, the classical protocols obtained high mDNA yields from sorghum tissues; however, they were less reproducible than the commercial kits. Commercial kits retrieved higher quality mDNA, but with lower endophytic bacterial diversities compared to classical protocols. The SDS-based protocol enabled access to the highest sorghum endophytic diversities. Therefore, “SDS-extracted” sorghum root and stem microbiome diversities were analysed via 454 pyrosequencing, and this revealed that the two tissues harbour significantly different endophytic communities. Nevertheless, both communities are dominated by agriculturally important genera such as Microbacterium, Agrobacterium, Sphingobacterium, Herbaspirillum, Erwinia, Pseudomonas and Stenotrophomonas; which have previously been shown to play a role in plant growth promotion. This study shows that DNA extraction protocols introduce biases in culture-independent studies of environmental microbial communities by influencing the mDNA quality, which impacts the microbial diversity analyses and evaluation. Using the broad-spectrum SDS-based DNA extraction protocol allows the recovery of the most diverse endophytic communities associated with sorghum tissues and, as such, establishes a reliable basis for future study of endophytic communities.

Keywords:
Metagenomic DNA extraction
Endophytic bacteria
Sorghum root and stem
t-RFLP
Pyrosequencing

1. Introduction

Endophytic bacteria play a crucial role in plant health and development, as they have either beneficial or detrimental effects. Beneficial endophytic bacteria also known as plant growth promoting endophytic bacteria (PGPeB) – enhance the plant’s growth and/or its ability to withstand stress (Schenk et al., 2012). They achieve this directly through production of plant-growth inducing phytohormones, such as indole acetic acid (IAA), gibberellins and auxins (Dodd et al., 2010), or indirectly via their varied nutrient-liberating metabolic activities such as nitrogen (N2) fixation, phosphate (P) solubilisation and iron sequestration (Khan et al., 2009; Kraiser et al., 2011; Saha et al., 2012). PGPeB have also been implicated in enhancing the plant’s resistance against pathogens through direct induction of the plant’s defence system or production of antimicrobial compounds (Heydari and Pessarakli, 2010). Pathogenic and parasitic endophytic bacteria, on the other hand, cause plant disease and reduced fitness (Newton et al., 2010).

Studies of endophytic bacterial communities associated with grass species (Poacea family) represent a well-established and continuously growing field (Seghers et al., 2004; Sun et al., 2008; Pereira et al., 2011; Sessitsch et al., 2012; Magnani et al., 2013); however, in-depth analyses in grasses such as sorghum are not available. Sorghum (and other grasses) is an important staple food in most developing countries (Taylor, 2004; Babalola and Glick, 2012) and, therefore, it is continuously being studied to increase yield of crop and its resistance to environmental and biological stress (Kapanigowda et al., 2013). The characterisation of the associated endophytic bacterial communities could make valuable contribution in this regard, and an understanding of their establishment and symbiotic roles within the plant has important implications in agriculture as it could lead to novel approaches in managing crop development and health (Schenk et al., 2012). Furthermore, endophytic...
of agricultural applications (biofertilizers and biocontrol agents), production of industrial/medical bioproducts and/or for bioremediation processes (Ryan et al., 2007; Andrews et al., 2010).

One of the primary challenges faced in culture-independent studies of plant-associated microorganisms is the retrieval of good-quality metagenomic DNA (mDNA) that can allow reliable downstream analyses using PCR-based techniques (e.g. T-RFLP, DGGE, pyrosequencing) (Demekel and Jenkins, 2010). The mDNA extraction must (i) ensure lysis of all microbial cells, (ii) provide sufficient genomic material (Terrat et al., 2012), and (iii) efficiently remove plant-derived phyto-chemicals, which may result in PCR-inhibition (e.g. polysaccharides, polyphenolic compounds, and secondary metabolites) and enzymes (e.g. DNases, proteinases) (Wilson, 1997; Demekel and Jenkins, 2010). Lysis of all microbial cells is particularly critical since the genomic mate-rial retrieved will dictate the bacterial diversity observed (Séné et al., 2001; Terrat et al., 2012). Thus, this study evaluated the effects of seven DNA extraction protocols commonly used in endophytic bacterial community studies, to determine the sorghum root and stem community diversities. Two classical protocols (CTAB- or SDS-based) and five commercial kits were compared using 2 different weights of sorghum tissue. The 16S rRNA gene diversity of retrieved endophytic bacterial communities was assessed using terminal restriction fragment length polymorphism (t-RFLP), and the most diverse community determined (SDS-extracted) was further characterised by higher resolution 16S rRNA gene amplicon pyrosequencing.

2. Material and methods

2.1. Study site and sampling procedures

Sorghum (Sorghum bicolor L. (Moench)) plant samples were collected from the Agricultural Research Council (ARC) academic farm situated in Potchefstroom (52°04’02’ E02°05’55’84’, North West Province, South Africa) during the autumn season (April, 2012). At the time of sampling, sorghum plants were mature at 16 weeks old. This crop has been continuously planted on the field annually for four years. Water was primarily supplied via rainfall events. Prior to planting, the soil was fertilized with “3:2:1 (32) + ZN” at a rate of 150 kg N/ha. LAN 28 fertilizer was applied at a rate of 100 kg/ha when plants were at knee length, and the soil was also treated with the insecticide, Kombat® (Kombat, South Africa). The soil on this field was a mixture of clay, silt and fine sand.

Three mature and healthy plants were collected following a random sampling technique. Stems and roots were aseptically excised and placed in sterile bags. All samples were immediately placed on ice and transported to the Institute for Microbial Biotechnology and Metagenomics (IMBM, University of the Western Cape, South Africa), where they were stored at −80 °C prior to processing.

2.2. Plant tissue sterilisation and preparation for metagenomic DNA extraction

The plant organs were sterilised using a modified protocol designed by Mendes et al. (2007). The roots and stems were separately washed in autoclaved double-distilled water until all residual soil was removed from their surfaces. Plant organs were immersed in 500 mL 1× phos-phate buffer solution (PBS) for 1.5 h, shaking at 1 rcf. The samples were then sequentially washed by shaking in (i) 70% ethanol for 10 min, (ii) 2.5% sodium hypochlorite solution for 20 min and (iii) rinsed five times in autoclaved double-distilled water for 2 min. To con-firm sterility, 100 μL of the last wash water was plated on nutrient agar (NA) and R2A agar plates (Merck, Germany) and incubated at room temperature for three days. Sterility tests were conducted in triplicate. The plant tissues were stored, unshaken, in the last wash water at 4 °C during these 3 days.

Sterilisation was considered successful when no colonies were observed on either plate. If microbial growth was observed, the sterilisation process was repeated. Sterilisation was repeated at least once per sample in this study. Sterilised root or stem samples were pooled and aseptically ground to a fine powder in liquid nitrogen using autoclaved pestle and mortar. Ground tissue powder aliquots (100 mg) were then stored at −80 °C.

2.3. DNA extraction procedures

DNA was extracted from 0.1 g to 0.3 g ground root and stem tissues. Seven plant mDNA extraction protocols were tested; two classical protocols (SDS- or CTAB-based) and five commercial kits (MoBio PowerPlant Pro® DNA Isolation Kit, Qiagen DNeasy® Plant Mini Kit, Fermentas GeneJet Plant Genomic DNA Purification Kit, MoBio PowerSoil™ DNA Purification Kit and MoBio UltraClean® Soil DNA Isolation Kit). These protocols have previously been used to study endophytic bacterial communities (Drabkova et al., 2002; Green et al., 1999; Krechel et al., 2002; West et al., 2010). All kit-based DNA extrac-tion protocols were performed according to the manufacturer’s instruc-tions, with the exception that starting plant material quantities were always 0.1 g or 0.3 g, and the final elution was performed in 50 μL buffer for normalization purposes. All DNA extractions were carried out in triplicate.

The CTAB-based mDNA extraction was a modified version of a proto-col described by Murray and Thompson (1980). B r i e f l y, 700 μL of a 1-bond strength CTAB buffer (100 mM Tris–HCl: 1.2 M NaCl; 20 mM EDTA; 2%CTAB; 0.2% β-mercaptoethanol) was added to ground plant tissue. The mixture was vortexed for 20 s (maximum speed) and incubated at 65 °C for 1 h. Subsequently, 500 μL isopropanol and RNase A (50 μg·mL−1; final concentration) were added to the super-natant in a clean tube and mixed by inversion, followed by incubation at room temperature for 20 min and centrifuged (13,000 rcf, 5 min). The supernatant was discarded, and the pellets were air dried. DNA pellets were washed twice with 250 μL 70% ethanol, which was eluted follow-ing centrifugation (13,000 rcf, 5 min). The DNA pellets were air dried and then resuspended in 50 μL TE buffer (10 mM Tris–HCl; 1 mM EDTA) before storage at −20 °C.

The SDS-based method used in this study is a modified version of the protocol developed by Zhou et al. (1996). Five hundred microlitres of lysozyme buffer (25 mM Tris–HCl; 50 mM glucose; 10 mM EDTA; 25 mg·mL−1 lysozyme) and RNase A (50 μg·mL−1; final concentration) were added to ground plant tissue powders and vortexed for 20 s. The mixtures were incubated at 37 °C for 1 h, and then treated with Proteinase K (1 mg·mL−1; final concentration) at 37 °C for 1 h. SDS was added to 1% final concentration, and mixed by inverting the tubes 10 times. Mixtures were incubated at 65 °C for 30 min. Tubes were centrifuged (14,000 rcf, 2 min) and the supernatants collected into new tubes. Equal volume phenol was added to each tube and mixed by inversion. Top aqueous phase containing DNA was collected after centrifugation (10,000 rcf, 1 min) and the bottom layer with organic phenol was discarded. The phenol extraction was repeated once. Equal volume 24:1 (v/v) chloroform/isoamyl alcohol solution was added to each tube and mixed by inversion. Top aqueous layer was collected and transferred to a new tube after centrifugation (10,000 rcf, 10 min). The tubes were placed on ice and equal volume ice-cold isopropanol was added, followed by incubation at 4 °C for 20 min. The tubes were centrifuged (10,000 rcf, 5 min) to recover mDNA and the isopropanol was discarded. DNA pellets were air dried under the laminar flow cabinet and then washed twice with 250 μL 70% ethanol, which was eluted after centrifugation (10,000 rcf, 5 min). The DNA pellets were air dried and then resuspended in 50 μL a t a r o m T E b u f fer a t a −20 °C.
2.4. PCR amplification and terminal restriction fragment length polymorphism (t-RFLP) analyses

Bacterial 16S rRNA gene amplification was conducted on DNA with starting materials of 10 ng, 5 ng and 1 ng per reaction, using the universal primers E9F (5′-GAGTTGTATCCTGGCTCAG-3′) and U1510r (5′-GGTACCTGGTTACGACTT-3′) (Marchesi et al., 1998; Turner et al., 1999). PCRs were carried out in 50 μl volumes containing 1× DreamTaq Buffer, 0.2 mM each dNTP, 0.5 M each primer, template DNA, 0.3 μl DreamTaq DNA polymerase (Fermentas, Lithuania) and deionised nuclease-free water, in a Labnet MultiGene™ Gradient PCR Thermal Cycler (Labnet International, Inc.). Cycling conditions were as follows: 4 min at 94 °C for initial denaturation; 30 cycles of 30 s denaturation at 94 °C, 30 s annealing at 52 °C and 105 s extension at 72 °C; and a final elongation step of 10 min at 72 °C. All PCRs were carried out in triplicate.

For t-RFLP analyses, the forward E9F primer (E9F) was 5′ end labelled with fluorescent dye fluorescein amide (FAM). The labelled PCR amplicons were purified with the Illustra GFX™ PCR DNA and Gel Purification Kit (GE Healthcare, UK) according to manufacturer's instructions. Purified PCR products (200 ng) were digested with the restriction enzyme Hae III (Fermentas, Lithuania) at 37 °C overnight. The lengths of fluorescently labelled terminally restricted fragments (t-RFs) were determined on an Applied Biosystems Genetic Analyzer sequencer (Applied Biosystems, Foster City, California, USA) at the Central Analytical Facility of the University of Stellenbosch, using the internal size standard marker, ROX1.1 (sizes in bp: 75, 100, 139, 150, 160, 200, 300, 340, 350, 400, 450, 500, 583, 683, 782, 932, 991, 1121). T-RFLP patterns were analysed using the Peak Scanner™ Software V1.0 (Applied Biosystems). Valid peaks (between 35 and 1000 bp) were identified and analysed using the online T-REX software (http://trex.biohpc.org/) (Culman et al., 2009). T-RFs were characterised by peak height and aligned to create an operational taxonomic unit (OTU) data matrix. The term OTU refers to an individual t-RF, with recognition that one t-RF could comprise more than one ribotype (Blackwood et al., 2007).

2.5. Pyrosequencing

The bacterial 16S rRNA gene was amplified from 0.3 g sorghum root and stem DNA extracted with the SDS-based protocol using multiplexed 8F (5′-CATCTCTCATCCCTGGTCTCTCAG-3′) and 518R (ATTACCCGCGCTGCTGG) primers (Turner et al., 1999; Muyzer et al., 1993). PCRs were carried out in 25 μl volumes containing 1× Phusion HF Buffer, 200 μM each dNTP, 0.5 μM each primer, 50 ng template DNA, 0.02 U/μl Phusion High Fidelity DNA Polymerase (Fermentas, Lithuania) and deionised nuclease-free water. PCR conditions were as follows: 2 min at 95 °C for initial denaturation; 25 cycles of 20 s denaturation at 98 °C, 15 s annealing at 75 °C and 90 s extension at 72 °C; 10 min extension at 72 °C. One composite mDNA sample was used per tissue. Five PCRs were prepared for each mDNA sample, and PCR products were pooled into single composite samples per tissue during the purification process and quantified. The amplicons were submitted to the Next Generation Sequencing Facility at the University of the Western Cape for pyrosequencing on the Roche 454 GS Junior platform (Roche, Branford, Germany).

2.6. Data analysis

Statistical tests for DNA yield and quality as well as diversity indices were performed using the software Sigma-Plot, Version 11.0 (Systat Software, Inc.). Two-way analysis of variance (ANOVA) was used to compare yield and purity of DNA extracted using different DNA extraction protocols. Normality tests were performed on the data following the method of Kolmogorov–Smirnov, with Lillifor’s correction (Justel et al., 1997). Data that did not pass the normality test was compared using the Holm–Sidak test that ranks the ordinal numbers and compares the median of the samples (Holm, 1979). Paired T-tests were conducted to compare differences in DNA yield at 0.1 g and 0.3 g starting plant material within individual tissues. Where normality criteria were not met, the Wilcoxon Signed Rank test was used (Wilcoxon, 1945).

The sample data matrix generated from t-RFLP profiles was analysed with the software Primer 6, version 6.1.11 (Primer E, Plymouth, UK). Diversity indices, i.e. species richness (S), Shannon index (H′) and the Simpson index (1−λ) (Clarke and Warwick, 2001), were calculated from untransformed data using the Diverse function. S is a direct count of operational taxonomic units (OTUs) observed. H′ measures the proportion of all OTUs in the whole community, and it is calculated as 

\[ H' = -\sum p_i \log(p_i) \]

where \( p_i \) is the proportion of the total count arising from the ith OTU. (1−λ) measures community evenness (or equitability), which quantifies how evenly distributed OTUs are within a community. It is calculated as

\[ 1 - \lambda = 1 - \frac{\sum N_i (N_i - 1)}{N(N - 1)} \]

where \( N_i \) is the number of OTUs that belong to species \( i \).

The T-RFLP dataset was standardised by the presence–absence transformation, and used to calculate the Bray–Curtis similarity coefficients between samples (Bray and Curtis, 1957), which were used to create a resemblance matrix. The latter in fine led to the construction of 3D-Nonmetric Multi-Dimensional (nMDS) plots, which are ordination of sample communities based on their relative similarities, i.e. the distance between two points reflects the degree of similarity between microbial community profiles (Clarke and Warwick, 2001; Shephard, 1962). Analysis of similarity (ANOSIM) was performed on the resemblance matrices to test for differences in structure, between a priori defined communities.

Pyrosequencing outputs from the 454 GS Junior software included a quality file (QUAL format) with information on the sequencing process and a metadata file (FASTA format) containing raw 16S RNA sequences. The two files were processed using the CloVR-16S pipeline version 1.1, which comprises of a suite of phylogenetic tools (Angiuoli et al., 2011). Preprocessing, processing and analysis of data were done using modules in QIIME (Quantitative Insights into Microbial Ecology http://qiime.org), R (http://www-R-project.org/) and MOTHUR (Schloss et al., 2009).

Quality assessment and filtering of raw sequences were performed using Prinseq (Schneider and Edwards, 2011). Multiplexed reads were split and assigned to samples based on their MID sequences using a Python script. Sequences were then trimmed and filtered to include only good-quality sequences of 200–470 bp. Sequences with high ambiguous base (N) occurrence and poly-A/T tails were removed. De novo chimera detection and OTU picking were performed with UCHIME and UCLUST, respectively (Edgar, 2010; Edgar et al., 2011). In this study, an OTU is defined as a cluster of sequences (Floyd et al., 2002), delimited at ≥97% sequence similarity. The representative sequences were aligned and used to create an OTU distance matrix with PyNAST (Caporaso et al., 2010). The distance matrix was converted to a phylogenetic tree with FastTree (Price et al., 2009). Taxonomy was assigned based on the Greengenes taxonomy and a Greengenes reference database (version 12.10) (McDonald et al., 2012), using the RDP Classifier (version 2.2) (Werner et al., 2012).

The OTU table was used to calculate the alpha diversity (within sample diversity). In this study, the Chao1 diversity index was calculated as a measure of true species diversity using the formula,

\[ S_{obs} = S_{obs}^F(F^2/F_2) \]

where \( S_{obs} \) is the number of species observed, \( F_1 \) is the number of singletons (occur once) and \( F_2 \) is the number of doubletons (Gotelli and Colwell, 2011). The Simpson index (1 − λ) was calculated to measure community evenness.

The taxonomic predictions for each OTU were used to create heatmaps and bar charts to reflect the distribution of phylotypes within each sample. UniFrac was then used to determine whether there was a significant difference in the structure of endophytic bacterial communities retrieved from the two tissues. Briefly, UniFrac measures the distance between each pair of environments as a fraction of the total
branch length in a phylogenetic tree, leading to sequences of one environment (Lozupone and Knight, 2005). To compare pairs of environments, the UniFrac value is calculated for each pair and a distance matrix is created (Lozupone and Knight, 2005; Lozupone et al., 2006, 2007). The two communities were then subjected to a statistical UniFrac test (F-test) to test for significance of difference at 95% confidence level, using the unweighted UniFrac metric for qualitative comparisons of β-diversities not affected by individual sequence abundance.

2.7. Characterisation of sorghum endophytic communities

The metabolic potential of sorghum endophytic communities was determined through literature survey. Properties of representative species of identified genera were recorded from published studies, in order to determine the agricultural and industrial potential of the sorghum endophyte. Sorghum endophytic communities analysed here were retrieved from unreplicated, but composite samples. Therefore, to exercise caution, only “dominant” phylotypes were characterised. In this case, “dominant” phylotypes constitute ≥0.1% of the total community, and these are most likely to be observed again in replicated studies (Charlson et al., 2012).

3. Results

3.1. Effects of DNA extraction protocols on mDNA yield and quality

Independently from starting plant material (0.1 g or 0.3 g) and plant tissue (stem or root), the classical CTAB- (ranging from 11.6 ± 2.8 to 19.1 ± 0.6 ng·mg⁻¹) and SDS-based (ranging from 48.0 ± 8.6 to 76.4 ± 4.3 ng·mg⁻¹) DNA extraction protocols retrieved significantly higher yields of sorghum mDNA than commercial kits (all ≤6.0 ± 0.3 ng·mg⁻¹) (ANOVA, p < 0.001) (Table 1), with also higher molecular weight (ranging between 10 kb and 20 kb). Appendices A1 and A2 show the visualisation of extracted mDNA by electrophoresis. The SDS-based protocol was particularly efficient when using 0.1 g starting plant material as it provided the highest mDNA yields for both stem and root tissue, with 48.0 (±8.5) ng·mg⁻¹ and 53.2 (±17.7) ng·mg⁻¹, respectively. However, when compared to the other protocols, it was the least reproducible with respect to mDNA yields (Table 1). Contrastingly, the other protocols were highly reproducible, despite significantly lower yields. Increasing the starting plant material to 0.3 g resulted in a significant increase in mDNA yield and reproducibility only for SDS-extracted sorghum stem mDNA, to 76.4 (±4.3) ng·mg⁻¹. The SDS-based protocol did not retrieve the most pure sorghum tissue mDNA (Table 1). In contrast, the CTAB-based protocol and Fermentas GeneJet kit consistently retrieved pure mDNA, whilst the Qiagen DNeasy and MoBio PowerPlant kits generated inconsistent results (i.e., dependent on either the starting material size or the tissue extracted). Soil DNA extraction kits (MoBio PowerSoil and MoBio UltraClean) were only used on 0.1 g starting plant material as both failed to extract high-yield and pure mDNA (Table 1).

Successful 16S rRNA gene PCR amplification from kit-extracted mDNA was consistently observed (as shown in Appendices B1 and B2), indicating that minimal PCR-inhibiting compounds were co-extracted. Contrastingly, PCR inhibition was more frequently observed when CTAB- and SDS-extracted mDNA was used as template. However, there was no clear correlation between PCR inhibition and template mDNA concentration used.

3.2. Effects of mDNA extraction protocols on endophytic bacterial community diversity and structure

T-RFLP was used to compare the endophytic bacteria community diversities associated with sorghum tissues. Since the MoBio PowerPlant and MoBio PowerSoil kits retrieved significantly low mDNA yields, they were excluded from these analyses.

With 0.1 g starting material, the number of OTUs (S) retrieved with the classical protocols from sorghum stem and root [8 (±1) to 10 (±3)] was significantly higher than those retrieved with commercial kits [3 (±1) to 7 (±3)] (ANOVA, p-values < 0.05) (Fig. 1). With Shannon index values (H′) of less than 1.7 and Simpson index (1 − λ) values ranging from 0.37 to 0.75, all sorghum endophytic communities retrieved had low diversity with low to moderate evenness.

Due to the consistent similarity in performance (mDNA yield and quality) observed between kit protocols (Table 1) in this study, only the Fermentas GeneJet kit was compared to classical protocols at 0.3 g starting plant material. An increase in starting plant material only resulted in a significant increase in community diversity with the SDS protocol (Fig. 1), with up to 31 (±3.0) and 23 (±3.0) OTUs retrieved from sorghum stem and root tissues, respectively. Contrastingly, CTAB-retrieved communities displayed significantly decreased richness and evenness when starting plant material was increased, whilst no changes in species richness or community evenness were observed when using the GeneJet kit (Fig. 1).

The 3D-Nonmorphic Multi-Dimensional Scaling (nMDS) ordinations of the community t-RFLP profiles retrieved from the “0.1 g and 0.3 g communities” presented stress values of 0.11 and 0.09, respectively, indicating low levels of scatter and minimal prospects of misinterpretation (Fig. 2) (Clarke and Warwick, 2001). Using 0.1 g plant material, nMDS ordination revealed two significantly different clusters (ANOSIM, Table 1

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Sorghum stem</th>
<th>Sorghum root</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.1 g Yield</td>
<td>0.3 g Yield</td>
</tr>
<tr>
<td>SDS</td>
<td>48.0 ± 8.66²</td>
<td>1.6 ± 0.0</td>
</tr>
<tr>
<td>CTAB</td>
<td>13.7 ± 1.17²</td>
<td>1.75 ± 0.0</td>
</tr>
<tr>
<td>GeneJet</td>
<td>4.4 ± 0.61²</td>
<td>2.3 ± 0.24²</td>
</tr>
<tr>
<td>DNeasy</td>
<td>5.93 ± 0.3d²</td>
<td>1.62 ± 0.0²</td>
</tr>
<tr>
<td>PowerPlant</td>
<td>4.1 ± 0.5d²</td>
<td>1.79 ± 0.0³</td>
</tr>
<tr>
<td>PowerSoil</td>
<td>1.9 ± 0.1e²</td>
<td>1.60 ± 0.0²</td>
</tr>
<tr>
<td>UltraClean</td>
<td>1.4 ± 0.0²d</td>
<td>1.57 ± 0.1²</td>
</tr>
<tr>
<td></td>
<td>53.2 ± 17.7³</td>
<td>1.55 ± 0.1</td>
</tr>
<tr>
<td></td>
<td>19.1 ± 0.6³</td>
<td>1.77 ± 0.0²</td>
</tr>
<tr>
<td></td>
<td>6.0 ± 0.3³³</td>
<td>1.77 ± 0.0²</td>
</tr>
<tr>
<td></td>
<td>4.9 ± 0.2³³</td>
<td>1.78 ± 0.0²</td>
</tr>
<tr>
<td></td>
<td>5.9 ± 1.0³³</td>
<td>2.01 ± 0.0³³</td>
</tr>
<tr>
<td></td>
<td>6.0 ± 0.2³³</td>
<td>2.5 ± 0.0³³</td>
</tr>
<tr>
<td></td>
<td>2.1 ± 0.6³³</td>
<td>1.62 ± 0.1³³</td>
</tr>
</tbody>
</table>

Values are given as mean ± standard deviation. Superscripts denote statistical comparisons in DNA yields (a-d for sorghum stem, and e-g for sorghum roots) (ANOVA, 95% significance level). DNA yield values with the same superscript letter are significantly similar. ND: not determined, as excluded from experiment.
Global R = 0.752, p-value = 0.001), clearly distinguishing sorghum root and stem communities, independently from the mDNA extraction protocol used (Fig. 2A). However, the structure of sorghum bacterial communities was also significantly affected by the DNA extraction protocol used (ANOSIM, Global R = 0.324; p-value = 0.001; Table 2). Communities retrieved with the SDS-based protocol from either tissue differed significantly from those retrieved with the other protocols (ANOSIM, 0.324 ≤ R ≤ 0.769; 0.01 ≤ p-values ≤ 0.05). The CTAB-protocol-retrieved communities were also significantly different to communities retrieved with commercial kits. When comparing commercial kits, a significant difference was only observed between communities retrieved with the Fermentas GeneJet kit and those with the MoBio UltraClean kit.

When starting plant material was increased to 0.3 g, sorghum root and stem communities remained significantly different (ANOSIM, Global R = 0.796; p-value = 0.002; Table 2), and community structures were even more significantly affected by the extraction protocol used (ANOSIM, Global R = 0.689; p-value = 0.001) (Fig. 2B, Table 2). SDS-retrieved communities were significantly different from CTAB- and GeneJet-retrieved communities, which were not significantly different (Table 2).

Interestingly, a generally strong clustering of the replicate communities from 3 different SDS-extracted mDNAs from either sorghum stem or root tissues is observed (Fig. 2B). This indicates that highly similar endophytic communities (structure and diversity) are reproducibly retrieved when using this protocol.

3.3. Diversity of sorghum endophytic communities by pyrosequencing

The most diverse sorghum-associated bacterial communities (based on microbial community fingerprinting) were retrieved when sorghum mDNA was extracted with the SDS-protocol from 0.3 g starting plant material. Therefore, for a higher resolution analysis, bacterial 16S rRNA pyrosequencing of sorghum mDNA was performed using these mDNA extracts.

A total of 13086 and 5831 sequences were generated from sorghum stem and root samples, respectively. These were curated and clustered into 829 OTUs, of which 15.6% (129/829) was shared between the

Fig. 1. Diversity of sorghum associated endophytic communities determined by t-RFLP. Diversity indices were calculated from t-RFLP profiles of stem and root communities retrieved with classical protocols and commercial kits from 0.1 g or 0.3 g plant tissue. Small letters (a-d) denote statistical (dis)similarities between endophytic communities (ANOVA, 95% confidence).
Genera dominant only in the root tissue include Leuconostoc (0.1%), Rhizobium (1.0%) and Sphingobium (4.3%), whilst genera restricted to the stem tissue include Rhodococcus (1.5%), Lactococcus (6.7%) and Swaminathania (2.4%). Approximately 5% of the phylotypes found in each tissue and affiliated to the α-Proteobacteria class could not be classified at the genus level.

Table 3 highlights the metabolic potential of the dominant sorghum endophytes identified. Dominant sorghum-associated bacterial genera such as Pseudomonas, Erwinia, Stenotrophomonas, Agrobacterium, Bacillus and Paenibacillus have previously been associated with plant growth promoting attributes (e.g. nutrient acquisition, antimicrobial potential and phytohormone production) as well as pathogenicity (Table 3). There is little reported on the role of Janthinobacteria and Methyloversatilis previous endophytic analyses.

### 4. Discussion

DNA extraction constitutes one of the most critical components in any cultivation-independent study; particularly when determining microbial community diversity (Demeke and Jenkins, 2010). Accurate estimation of community profiles and diversity relies on the DNA’s representativeness of the indigenous communities, and its usability in PCR (Demeke and Jenkins, 2010; Terrat et al., 2012). The current study established that different plant mDNA extraction procedures lead to highly variable mDNA quantity and quality, which is reflected in the endophytic bacterial community diversities observed.

Conventional SDS- or CTAB-based protocols retrieved high yields of sorghum mDNA compared to commercial kit protocols, as previously observed on fresh and historical plant samples (Drabkova et al., 2002; Niu et al., 2008; Sahu et al., 2012). Metagenomic DNA yield relies on the efficiency of the cell lysis step, which includes mechanical (e.g. grinding, dead-mill) and chemical (e.g. enzymatic lysis) processes for cell disruption (Moré et al., 1994). Therefore, it is reasonable to conclude that the combination of tissue grinding under liquid nitrogen and lysis buffers used in classical protocols was more efficient than the lytic procedures of commercial kits used in this study. It is important to note that plant DNA extracting kit manufacturers recommend the use of a bead-mill for increased mDNA yield, and its efficiency in this regard was previously demonstrated (Miller et al., 1999). However,

### Table 2

<table>
<thead>
<tr>
<th>Differences among DNA extraction protocols</th>
<th>0.1 g plant material</th>
<th>0.3 g plant material</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global Test</td>
<td>0.324</td>
<td>0.287</td>
</tr>
<tr>
<td>CTAB vs SDS</td>
<td>0.454</td>
<td>0.02*</td>
</tr>
<tr>
<td>CTAB vs GeneJet</td>
<td>0.287</td>
<td>0.02*</td>
</tr>
<tr>
<td>CTAB vs DNeasy</td>
<td>0.454</td>
<td>0.02*</td>
</tr>
<tr>
<td>CTAB vs UltraClean</td>
<td>0.352</td>
<td>0.11</td>
</tr>
<tr>
<td>SDS vs GeneJet</td>
<td>0.769</td>
<td>0.01*</td>
</tr>
<tr>
<td>SDS vs DNeasy</td>
<td>0.324</td>
<td>0.05*</td>
</tr>
<tr>
<td>SDS vs UltraClean</td>
<td>0.352</td>
<td>0.05*</td>
</tr>
<tr>
<td>GeneJet vs DNeasy</td>
<td>0.139</td>
<td>0.36</td>
</tr>
<tr>
<td>GeneJet vs UltraClean</td>
<td>0.287</td>
<td>0.03*</td>
</tr>
<tr>
<td>DNeasy vs UltraClean</td>
<td>-0.028</td>
<td>0.54</td>
</tr>
</tbody>
</table>

R: ANOSIM Statistic; p: probability level.

* Statistically different (p < 0.05).
Fig. 3. Relative abundance of major bacterial lineages recovered from composite samples of sorghum root and stem tissues. Relative abundance is calculated as the percentage of sequences belonging to a particular lineage out of all 16S rRNA gene sequences recovered from each plant tissue.

Fig. 4. Venn diagram representation of the composition of sorghum endophytic bacterial communities in stem and root tissues. A total of 134 phylotypes were identified from sorghum tissues through 16S rRNA gene amplicon pyrosequencing, with 50 phylotypes shared between the root and stem tissues. Twenty-one "dominant" phylotypes were observed (underlined, * indicates phylotypes accounting for over 0.1%, ** for over 5% and *** for over 10% of the total population in the root and/or stem tissue).
for the benefit of consistency in comparisons, this was not used in this study.

Despite lower DNA yields, commercial kits – notably, the Fermentas GeneJet kit – extracted mDNA of superior PCR-quality compared to classical protocols (Green et al., 1999; Drábková et al., 2002). The observed low purity of mDNA retrieved with classical protocols and the low PCR-efficiency when it was used as template in PCR could result from a high concentration of co-extracted plant polyphenolics and polysaccharides which are known to bind to DNA, making it inaccessible to the polymerase enzyme (Horne et al., 2004; Varma et al., 2007; Deméke and Jenkins, 2010; Mornkham et al., 2012). Indeed, it was previously shown that chloroform and phenol in classical protocols were not consistent in efficiently removing potential PCR inhibitors. Moreover, reagents used in these processes such as CTAB, SDS, phenol, chloroform and ethanol, can also contaminate extracted mDNA and lead to PCR inhibition (Deméke and Jenkins, 2010).

T-RFLP was effective in evaluating the effects of plant mDNA extraction protocols. Classical protocols enabled access to more diverse endophytic bacterial communities than commercial kits. It is unclear from these findings, whether there is a relationship between plant mDNA yield and the diversity of endophytic bacteria retrieved. However, Scupham et al. (2007) and Salonen et al. (2010) have argued that DNA yield or purity was not always correlated to the diversity of communities retrieved, but that it was directly influenced by the lysis step of an extraction protocol. Robust lytic processes (e.g. SDS-treatment) are able to rupture a broader range of bacterial cells, including hard cell-walled and endospore-forming bacteria such as Firmicutes and Actinobacteria (Moré et al., 1994; Yuan et al., 2012).

The effects of plant material size on DNA recovery and endophytic community diversity varied with the extraction protocols used. With the SDS-protocol, mDNA yield and bacterial community diversity improved with an increase in plant tissue material, whilst the CTAB protocol and commercial kits had a limiting or no significant effect on mDNA yield or community diversity. To our knowledge, there is no published study that has assessed the effects of plant sample size on mDNA yield and bacterial community structure. However, previous studies on soil samples have indicated that correlations between sample size and DNA recovery or bacterial diversity are dependent on the type of soil

Table 3

<table>
<thead>
<tr>
<th>Phylum/class</th>
<th>Genus</th>
<th>H</th>
<th>N</th>
<th>P</th>
<th>S</th>
<th>X</th>
<th>M</th>
<th>A</th>
<th>C</th>
<th>D</th>
<th>Root Stem</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actinobacteria</td>
<td>Curtobacterium</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>1.6</td>
<td>(4)</td>
<td></td>
<td></td>
<td>2.0 (7)</td>
<td>Aizawa et al. (2007), Rajkumar et al. (2009), Sowmya et al. (2013), Chen and Yin (2007)</td>
</tr>
<tr>
<td></td>
<td>Microbacterium</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>0.7</td>
<td>(8)</td>
<td></td>
<td></td>
<td>0.2 (3)</td>
<td>Tsavkelova et al. (2006), Rajkumar et al. (2009), Lal et al. (2008)</td>
</tr>
<tr>
<td></td>
<td>Rhodococcus</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>0.0</td>
<td></td>
<td></td>
<td>1.5 (4)</td>
<td>Tsavkelova et al. (2006), Ye et al. (2004), Ambrosini et al. (2012), Rajkumar et al. (2009), Cornelis et al. (2001), Al-Awadhi et al. (2009), Chen et al. (2006), Li et al. (2012)</td>
</tr>
<tr>
<td>Bacteroidetes</td>
<td>Chryseobacterium</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>0.3</td>
<td>(2)</td>
<td></td>
<td></td>
<td>0.0 (1)</td>
<td>Marques et al. (2010), Radianingtyas et al. (2003), Ambrosini et al. (2012), Luo et al. (2011a), Lucas Garcia et al. (2004), Singh et al. (2013), Berg et al. (2005)</td>
</tr>
<tr>
<td>Firmicutes</td>
<td>Bacillus</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>1.2</td>
<td>(4)</td>
<td></td>
<td></td>
<td>0.0 (2)</td>
<td>Tsavkelova et al. (2006), Rodriguez and Fraga (1999), Ye et al. (2004), Rajkumar et al. (2009), Cho et al. (2007), Sahin et al. (2004), Berg et al. (2005)</td>
</tr>
<tr>
<td></td>
<td>Paenibacillus</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>1.8</td>
<td>(2)</td>
<td></td>
<td></td>
<td>0.1 (2)</td>
<td>Rajkumar et al. (2009), Cho et al. (2007), Daane et al. (2001), Lal and Tabacchioni (2009)</td>
</tr>
<tr>
<td></td>
<td>Leuconostoc</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>0.1</td>
<td>(1)</td>
<td></td>
<td></td>
<td>0.0</td>
<td>Harding and Shaw (1990), Gillespie et al. (2002)</td>
</tr>
<tr>
<td></td>
<td>Lactococcus</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
<td>0.0</td>
<td></td>
<td></td>
<td>6.7 (7)</td>
<td>Barré et al. (2007), Rodriguez et al. (2005)</td>
</tr>
<tr>
<td>Alphaproteobacteria</td>
<td>Unclassified</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
<td>5.3</td>
<td>(15)</td>
<td></td>
<td>5.2 (15)</td>
<td>Tsavkelova et al. (2006), Rodriguez and Fraga (1999), Ye et al. (2004), Rajkumar et al. (2009), Cho et al. (2007), Sahin et al. (2004), Berg et al. (2005)</td>
</tr>
<tr>
<td></td>
<td>Agrobacterium</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
<td>8.7</td>
<td>(4)</td>
<td></td>
<td>14.2 (4)</td>
<td>Tsavkelova et al. (2006), Rodriguez and Fraga (1999), Ye et al. (2004), Rajkumar et al. (2009), Cho et al. (2007), Sahin et al. (2004), Berg et al. (2005)</td>
</tr>
<tr>
<td>Betaproteobacteria</td>
<td>Herbaspirillum</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>1.0</td>
<td>(1)</td>
<td></td>
<td></td>
<td>1.0</td>
<td>Tsavkelova et al. (2006), Franche et al. (2009), Rodriguez and Fraga (1999), Ambrosini et al. (2012), Wani et al. (2008)</td>
</tr>
<tr>
<td></td>
<td>Swaminathania</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
<td>0.0</td>
<td></td>
<td>2.4</td>
<td>(9)</td>
<td></td>
<td>Loganganth and Nair (2004)</td>
</tr>
<tr>
<td></td>
<td>Sphingobium</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
<td>4.3</td>
<td>(1)</td>
<td></td>
<td>0.0</td>
<td>Raina et al. (2008), Ambrosini et al. (2012), Pavić et al. (2011), Wang et al. (2013), Pereira and Castro (2014)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sphingomonas</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>0.3</td>
<td>(2)</td>
<td></td>
<td>2.7</td>
<td>(7)</td>
<td>Ye et al. (2004), Rajkumar et al. (2009), Franche et al. (2009), Berg et al. (2005), Buonaurio et al. (2002), Romanenko et al. (2007)</td>
</tr>
<tr>
<td>Gammaproteobacteria</td>
<td>Janthinobacterium</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
<td>0.3</td>
<td>(2)</td>
<td></td>
<td>0.1</td>
<td>(1)</td>
<td>Huang et al. (2012), Bull et al. (2010)</td>
</tr>
<tr>
<td></td>
<td>Methyloversatilis</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
<td>0.1</td>
<td>(3)</td>
<td></td>
<td>0.4</td>
<td>(4)</td>
<td>Tsavkelova et al. (2006), Ambrosini et al. (2012), McGuinness and Dowling (2009), Berg et al. (2005), Bull et al. (2010), Ting et al. (2011)</td>
</tr>
<tr>
<td></td>
<td>Erwinia</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
<td>0.1</td>
<td>(5)</td>
<td></td>
<td>30.2</td>
<td>15</td>
<td>Tsavkelova et al. (2006), Rodriguez and Fraga (1999), Mansfield et al. (2012), Hiraishi (2003), Boddey et al. (2003), Tian et al. (2009)</td>
</tr>
<tr>
<td></td>
<td>Pseudomonas</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
<td>46.5</td>
<td>(34)</td>
<td>10.1</td>
<td>29</td>
<td>Tsavkelova et al. (2006), Miller et al. (2010), Radianingtyas et al. (2003), Ambrosini et al. (2012), Cho et al. (2007), Mansfield et al. (2012), Berg et al. (2005)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stenotrophomonas</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
<td>4.7</td>
<td>(10)</td>
<td>11.5</td>
<td>23</td>
<td>Tsavkelova et al. (2006), Radianingtyas et al. (2003), Ambrosini et al. (2012), Rajkumar et al. (2009), Xiao et al. (2009), Berg et al. (2005), Wolf et al. (2002), Suckstorff and Berg (2003)</td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>89.4</td>
<td>89.9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
used (Ranjard et al., 2003; Kang and Mills, 2006; Zhao and Xu, 2012; Leite et al., 2014). Therefore, we recommend investigating the sample size factor prior to any extensive environmental microbial community analysis. Furthermore, Ranjard et al. (2003) pointed out that whilst larger samples improve the reproducibility of fingerprinting analyses, smaller samples (between 0.1 g and 1 g) such as those used in the current study are able to reveal rare phylogenotypes that are otherwise obscured by dominant phylogenotypes when larger samples are used.

The use of the different protocols led to significantly different sorghum-associated bacterial communities. This indicates that the DNA extraction procedure introduces significant bias, thereby influencing the interpretation of the microbial community observed. PCR-based metagenomic approaches such as DGGE (Hardoim et al., 2012; Ramond et al., 2013), t-RFLP (Sessitsch et al., 2012; Ding et al., 2013) and next generation sequencing (Gottel et al., 2011; Lucero et al., 2011; Inceoglu et al., 2011; Molina et al., 2012) are continuously used to elucidate the structure and ecological roles of plant-associated (and other) microbial communities. However, few, if any, of these studies share the same DNA extraction protocol and/or are preceded by an evaluation of DNA extraction procedures to select the most efficient. Therefore, discrepancies in community profiles retrieved by different procedures pose a challenge in the reproducibility of previous findings as well as comparisons between studies.

In the current study, the SDS-protocol was established as the most efficient protocol in retrieving diverse endophytic bacterial communities from sorghum tissues. This protocol was further used to extract sorghum root and stem mDNA to perform high-throughput pyrosequencing using 16S rRNA marker analysis. Sorghum-associated bacterial communities were shown to be highly diverse, and root and stem communities were significantly different (Castro-Carrera et al., 2014; Lagos et al., 2014).

To date, the diversity of the sorghum endophyte remains largely unexplored. In one study, bacterial isolates from sorghum tissues were affiliated to the phyla Firmicutes, Actinobacteria, α-Proteobacteria and γ-Proteobacteria, with no further characterisation (Grönemeyer et al., 2012). Ramond et al. (2013) showed that the diversity of sorghum associated endophytic bacteria is lower than that of rhizospheric communities using t-RFLP and DGGE; however, the low resolution of the two fingerprinting techniques made it difficult to assign taxonomy for the recovered OTUs. Despite this, sorghum has been used widely as a host organism in the study of PGPEs such as *Gluconacetobacter* species and *Bacillus* sp. SLS18, *Agrobacterium* sp. (Luna et al., 2010; Luo et al., 2011b; Wu et al., 2014) as well as pathogenic bacterial strains such as *Herbaspirillum rubrisubalbicans* (James et al., 1997).

The most dominant OTUs in this study belong to 20 different bacterial phylotypes, 19 of which were identified at genus level (Table 3). These genera have previously been detected/isolated from tissues of other graminaceous and non-graminaceous plants including maize, sugarcane, rice, poplar, grapevine and sunflower (Pereira et al., 2011; Magnani et al., 2010; Sun et al., 2008; Ulrich et al., 2008; Compant et al., 2011; Ambrosini et al., 2012), thus indicating that these genera are probably ubiquitous in endophytic environments. This study also shows that the structures of sorghum stem and root communities are significantly different, and this could be explained by the differences in the selectivity of recruitment processes of microbes into each niche environment (Hardoim et al., 2008). The root tissue is the primary entry point for bacteria as it is in contact with the rhizosphere (Lugtenberg and Kamilova, 2009). All bacteria in the vicinity of the root can be recruited into the root, and as such, the root endophyte is diverse (Hardoim et al., 2008). Only the most competitive endophytes can survive within the plant tissue environment, and a selection of these will be able to move within the plant to other plant tissues. As such, bacterial communities in aerial parts of the plant are often uniquely composed, and significantly less diverse (Hardoim et al., 2008). However, in this study, sorghum stem communities were slightly more diverse than sorghum root communities. This further supports the hypothesis that only a small proportion of sorghum-associated endophytic bacteria are specifically recruited to different niches, whereas the bulk of the endophytic community is composed of opportunistic endophytes that are subjected to minimal selective pressure (Ramond et al., 2013).

Dominant phylogenotypes in sorghum included ecologically and biotechnologically significant bacterial genera. Genera including *Pseudomonas*, *Erwinia*, *Stenotrophomonas*, *Paenibacillus*, *Bacillus* and *Herbaspirillum* have previously been shown to play important and varied roles in plant life (Table 3). For example, representatives of *Bacillus*, *Pseudomonas* and *Rhizobia* species have been shown to produce plant growth promoting hormones (Matiru and Dakota, 2004; Sahu et al., 2012; Reis et al., 2011) and also include numerous diazotrophic bacteria that are able to mineralise atmospheric nitrogen and/or solubilise phosphate compounds, to the benefit of the plant (Reis et al., 2011; Kraiser et al., 2011). Genera with potential plant-growth promoting properties can be targeted and isolated for development of biofertilizer or biocontrol products. However, these bacterial genera also include well-known plant pathogens, which could affect crop production and food security, such as *Pseudomonas syringae*, *Erwinia carotovora* and *Erwinia amylovora* (Bender et al., 1999; Basset et al., 2000; Oh and Beer, 2005; Mohr et al., 2008), as well as human pathogens including *Pseudomonas aeruginosa*, *Rhodococcus equi* and *Bacillus cereus* (Gován and Deretic, 1996; Weinstock and Brown, 2002; Bottone, 2010).

### 5. Conclusion

The current study showed that DNA extraction protocols introduce significant biases in the endophytic community diversities. Conventional protocols provide high-yield mDNA; however, the quality of extracted mDNA is low. The opposite is true for commercial kit protocols; they retrieve low yield mDNA of good quality for application in PCR. It can be concluded that the cell lysis procedures of classical protocols were superior to those of commercial kits; whilst the latter, had more efficient mDNA purification processes in comparison. By this argument, future research should, therefore, seek a feasible compromise between cell lysis and DNA purification, towards the development of a broad-spectrum DNA extraction procedure that retrieves PCR-ready metagenomic DNA, which is representative of endophytic communities.

Pyrosequencing analysis revealed that sorghum tissue harbours diverse bacterial phylotypes with the potential of being applied for the improvement of sorghum health and production. However, 16S rRNA analysis using pyrosequencing cannot adequately identify the diversity at the species level, which is the requirement in order to develop targeted species with plant growth promoting attributes. Therefore, future studies need to involve the identification and isolation of sorghum tissue specific endophytic species. This study provides the initial identification of bacteria recruited by sorghum tissues, and suggests that a strategy to develop sorghum specific applications is foreseeable.

### Conflict of interest

The authors declare that there are no competing interests.

### Acknowledgements

This study was wholly sponsored by the South African National Research Foundation (NRF) Grant No: 71081. However, the funders did not play any role in any other aspects of the research such as study design; data collection, analysis and interpretation; report writing and the decision-making process regarding publication of the study. Plant samples used in this study were obtained from the Agricultural Research Council (ARC), Potchefstroom, South Africa, where we were graciously hosted and assisted by Dr. Nemeria Sharige during sampling. The analysis of NGS data was carried out at the Institute of Infectious Disease and Molecular Medicine (IIDMM), at the University of Cape Town, and technical assistance was provided by Mr. Gerrit Botha.
Appendix A

Appendix A1. Sorghum mDNA extracted from 0.1 g ground tissue. DNA was extracted from stem and root tissues using classical protocols (CTAB- or SDS-based) and commercial kits (Fermentas GeneJet Plant Genomic DNA Purification Kit, Qiagen DNeasy® Plant Mini Kit, MoBio PowerPlant Pro® DNA Isolation Kit, MoBio PowerSoil™ DNA Purification Kit and MoBio UltraClean® Soil DNA Isolation Kit), and 5 μL visualised by electrophoresis (80 V, 1 h) on 0.8% agarose gels. DNA molecular size was determined by comparison to molecular markers, lambda DNA digested with HindIII (H) and PstI (P) restriction enzymes. Extractions were performed in triplicate.

Appendix A2. Sorghum mDNA extracted from 0.3 g ground tissue. DNA was extracted from stem and root tissues using classical protocols (CTAB- or SDS-based) and commercial kits (Fermentas GeneJet Plant Genomic DNA Purification Kit, Qiagen DNeasy® Plant Mini Kit and MoBio PowerPlant Pro® DNA Isolation Kit), and 5 μL visualised by electrophoresis (80 V, 1 h) on 0.8% agarose gels. DNA molecular size was determined by comparison to molecular marker, lambda DNA digested with HindIII (H) restriction enzyme. Extractions were performed in triplicate.
Appendix B1. Amplification of bacterial 16S rRNA gene from sorghum mDNA (from 0.1 g tissue). Bacterial 16S rRNA gene amplified from stem (SS1–SS3) and root (SR1–SR3) mDNA (triplicate samples) extracted using the CTAB-protocol (A), SDS-protocol (B) and commercial kits: Fermentas GeneJet Plant Genomic DNA Purification Kit (C), MoBio PowerPlant Pro® DNA Isolation Kit (D), Qiagen DNeasy® Plant Mini Kit (E), MoBio PowerSoil™ DNA Purification Kit (F) and MoBio UltraClean® Soil DNA Isolation Kit (G). Template mDNA was tested at 10 ng, 5 ng and 1 ng per reaction. *Geobacillus* sp. genomic DNA (5 ng per reaction) and autoclaved double-distilled water were used as positive (+) and negative (−) controls, respectively. PCR products (5 μL) were visualised via electrophoresis (80 V, 30 min) on 1.5% agarose gels, and size determined by comparison to lambda PstI marker.

Appendix B2. Amplification of bacterial 16S rRNA gene from sorghum mDNA (from 0.1 g tissue). Bacterial 16S rRNA gene amplified from stem (SS1–SS3) and root (SR1–SR3) mDNA (triplicate samples) extracted using the CTAB-protocol (A), SDS-protocol (B), Fermentas GeneJet Plant Genomic DNA Purification Kit (C), Qiagen DNeasy® Plant Mini Kit (D) and MoBio UltraClean® Soil DNA Isolation Kit (E). PCR-products were visualised via electrophoresis (80 V, 30 min) on 1.5% agarose gels, and size determined by comparison to the lambda PstI molecular marker.

Appendix C. Biom information of sequences retrieved from bacterial 16S rRNA pyrosequencing of sorghum mDNA processed with Qiime (http://qiime.org).

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Root</th>
<th>Stem</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of samples</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Number of sequences</td>
<td>5831.0</td>
<td>13086.0</td>
</tr>
<tr>
<td>Number of OTUs</td>
<td>331</td>
<td>498</td>
</tr>
</tbody>
</table>

Appendix D. True diversity (A) and evenness (B) of endophytic bacterial communities recovered from sorghum root and stem tissues.


