Improved real-time PCR estimation of gene copy number in soil extracts using an artificial reference

T.J. Daniella,⁎, J. Davidsona, C.J. Alexanderb, S. Caula, D.M. Robertsa

a Ecological Sciences, The James Hutton Institute, Invergowrie, Dundee, DD2 5DA, UK
b Biomathematics & Statistics Scotland, Invergowrie, Dundee, DD2 5DA, UK

Abstract
Application of polymerase chain reaction (PCR) techniques has developed significantly from a qualitative technology to include powerful quantitative technologies, including real-time PCR, which are regularly used for detection and quantification of nucleic acids in many settings, including community analysis where culture-based techniques are not suitable. Many applications of real-time PCR involve absolute quantification which is susceptible to inaccuracies caused by losses during DNA extraction or inhibition caused by co-extracted compounds. We present here an improvement to this approach involving the addition of an artificial internal standard, prior to nucleic acid extraction. The standard was generated by in-situ mutagenesis from an E. coli template to ensure it both did not amplify with bacterial primers used for quantification and was short enough to minimise possible interference with other analyses. By estimating gene target copies by relative abundance, this approach accounts for both loss during extraction and inhibition effects. We present a novel application of relative real-time PCR, using the internal standard as a reference, allowing accurate estimation of total bacterial populations both within and across a wide range of soils and demonstrate its improvement over absolute quantification by comparison of both approaches to ester linked fatty acid analysis of the same soils.

1. Introduction
Analysis of soil communities is affected by the huge observed levels of microbial diversity (Curtis et al., 2002) combined with low level of culturability (Torsvik and Øvreås, 2002) leading to the wide application of molecular techniques for the analysis of community structure. This often involves the application of techniques such as denaturing gradient gel electrophoresis (DGGE) or terminal restriction fragment length polymorphism (T-RFLP) which rely on polymerase chain reaction (PCR) amplification of diagnostic genomic regions to target groups either phylogenetically or functionally and produce data at a presence/absence or relative abundance level (for a review see Hirsch et al., 2011). These techniques are limited to the analysis of community structure and are incapable of estimating community size which has led to the application of real-time PCR methods using gene copy number as a proxy of community size (Hermansson and Lindgren, 2001; Hiroshi et al., 2011; Macdonald et al., 2011). Alternatively a wide range of techniques including lipid based methods such as phospholipid-derived fatty acid (PLFA) and ester linked fatty acid (ELFA) (Frostegård and Bååth, 1996), direct counting (Bogosian et al., 1996) or chloroform fumigation (Anderson and Domsch, 1978) have been used to estimate population size or biomass. Typically, application of real-time PCR methods in soil systems targets DNA in order to estimate target-gene copy as a proxy for population size, for example bacterial quantification using fragments of the 16S gene (Stubner, 2002), or quantifying genes to target functional groups such as the ammonium monooxygenase A subunit to estimate nitrifier population size (Okano et al., 2004). The application of real-time PCR and other molecular methods for the quantification of functional groups is reviewed in Sharma et al. (2007). The majority of applications have applied absolute methods where a standard curve is generated from a serial dilution of a clone of the target amplicon allowing estimation of copy number in reactions using the same conditions. Whilst population estimation using molecular approaches have many advantages over culture-based or microscopy-based methods, they do have limitations. They can, for example, suffer from inhibition of any PCR amplification by co-extracted compounds such as humic acids that inhibit enzymatic reactions, including PCR amplification (Jackson et al., 1997; Okubara et al., 2005; Zhou et al., 1996), or inadequacies in the DNA extraction that can lead to losses that are difficult to predict or quantify. Losses can be expected at every step of extraction including cell lysis and losses during purification, for instance where phenol:chloroform extraction and ethanol precipitation steps occur in the DNA extraction process (Mymy and Findlay, 2004; Tsai et al., 1991). For example, losses during ethanol precipitation of between 70% and 20% dependent on a number of factors including initial DNA levels, temperature and length of centrifugation (Zeugin and Hartley, 1985) and phenol:chloroform extraction efficiencies.
of 49–59% have been estimated for community extractions (Weinbauer et al., 2002). These potential issues make accurate estimation of copy number difficult with absolute real-time PCR quantification, especially if experiments include soil from different origins or with treatments likely to alter the level of inhibitors.

Many applications of real-time PCR, predominantly in relatively clean systems such as culture based experiments or large organism studies where samples are taken from single individuals, estimate expression levels of genes of interest using housekeeping genes to normalise results (Pfaffl et al., 2002). This option is not available to estimate gene copy number in soil systems since sample size, in terms of organism number or biomass, is hard if not impossible to standardise.

A similar approach has been suggested for molecular approaches including PCR from soil (using DGGE), for example Park and Crowley (2005) who used Escherichia coli JM 109 cells containing plasmid which were added to soil and detected using primers specific for the plasmid in real-time PCR analyses to normalise results. Additionally the presence of the E. coli band in DGGE was used to normalise intensity across this analysis. The issue here is that the presence of E. coli genomic DNA potentially skews estimates of any target contained within this genome, complicating analysis. The ideal situation would be to utilise a normalisation target and methodology that can both assess extraction and PCR efficiency whilst not interfering in any way with normal analysis of the samples. To achieve this any spike has to be independent of any intended analysis methodology which precludes the use of a natural DNA template.

Here we aimed to improve the accuracy of real-time PCR estimation of gene-copy number in DNA extracted from soil samples through the application of an artificial reference spike and relative real-time PCR. To achieve this aim we generated the spike through in-situ mutagenesis of an E. coli DNA template for amplification of a fragment of the 16S ribosomal gene altering the three 3′ terminal bases of the PCR primer target sequences to their complementary bases. This spike would thus be amplified by the normal primer set used to target the 16S gene or, due to its short length, primers normally used for community structure analysis allowing extracts to be used for normal analysis with no interference from the added reference spike. The mutated spike was then added to a wide range of soil types and DNA extracted and used to test the efficiency of relative real-time PCR targeting the bacterial 16S gene as a proxy for cell number or biomass using the spike as reference for relative quantification. We have further contrasted results against estimates generated using the same amplification but applying absolute quantification and comparing both with an established method to estimate soil biomass. The application of this methodology should allow far more accurate control of technical variability recently recognised by Bustin (2010) as a primary cause for inconsistency and irregularity in quantitative PCR results.

2. Methods

2.1. Soil collection/storage

The location of each sample site is given in Supplementary Table 1. These sites were selected to cover a wide range of soil types. Six cores were randomly taken at each site, to a depth of ten centimetres using a grass plot sampler (Van Walt Ltd, Haslemere, UK), and combined to obtain a homogenous sample from each site. Samples were sieved to 4 mm to remove large debris, mixed and eight replicate sub-samples taken. Four were used to allow the assessment of the accuracy of estimation of each soil type by providing measurements from each sub-sample using both real-time PCR techniques. The other set of four sub-samples were used to measure soil bacteria using the ELFA technique as a ‘gold standard’. Samples were snap-frozen in liquid nitrogen and aliquots stored at −80 °C prior to DNA or lipid extraction.

2.2. ELFA

Ester linked fatty acids (ELFA) were extracted as described by Schutter and Dick (2000). Briefly, frozen aliquots of soil were extracted with KOH in methanol, neutralised with acetic acid and fatty acid methyl esters extracted with hexane. Following drying, fatty acid methyl esters were re-dissolved in 100 μl isohexane and analysed by gas chromatography on a Agilent 6890N Network system fitted with a flame ionization detector, Agilent HP-5 column (Agilent Technologies UK Ltd, Stockport, UK). Hydrogen was used as carrier gas and helium as the make-up gas, with the following programme: 160 °C for 2 min, an increase by 4 °C per min to 270 °C, hold for 10 min. Fatty acid 23:0 was added as an internal standard. Double bonds of fatty acids were referred to the methyl end (ω) of the molecule and the nomenclature for ELFA was used as described previously (Zogg et al., 1997). Bacterial biomass was calculated as the concentration of fatty acids that could be related to bacterial groups (Zelles, 1999).

2.3. Preparation of calibration controls and of reference spike

The 16S wild type control was generated by PCR amplification of a 16S gene fragment from E. coli genomic DNA generated from a lysate generated by boiling a colony suspended in 100 μl water for 5 min. 1 μl of this preparation was then used as template for PCR using Primer-1 and Primer-2 (all primer sequences in Table 1) in a final reaction volume of 25 μl containing: 2.5 U DNA polymerase (Expand High Fidelity enzyme mix. Roche Diagnostic Systems, Burgess Hill, UK), 5 pmol of each primer, 0.5 mM of each nucleotide, 1× Expand High Fidelity buffer (Roche Diagnostic Systems, Burgess Hill, UK) and 10 μg of bovine serum albumin (BSA) (Roche Diagnostic Systems, Burgess Hill, UK). PCR was performed with a DNA Engine DYAD thermocycler (MJ Research, UK) with an initial denaturation step of 95 °C for 5 min followed by 30 cycles of 95 °C for 1 min, 34 °C for 1 min and 72 °C for 30 s; cycling was completed by a final extension period of 72 °C for 5 min.

The reference spike was generated by PCR in situ mutagenesis (Vallet et al., 1989) to form a product in which the three 3′ terminal recognition bases of both PRIMER-1 and PRIMER-2 were altered to the requisite complementary bases providing a template suitable for amplification with the Mut-F and Mut-R primers.

This was performed in a two stage PCR process using 342FMut and Primer 2 in the first reaction and Mut-F and 534RMut in the second using the PCR conditions given above. 1 μl E. coli genomic DNA was used as template in round one and 1 μl of 100 fold diluted rounded round in the second reaction. PCR success was assessed by 1.5% agarose gel electrophoresis.

Products were cloned into pGEM-T Easy (Promega, Southampton, UK), following the manufacturer’s instructions and transformed into E. coli DH10B electroporment cells prepared following the method of Tun and Chow (1995). Colonies were screened for mutation success by PCR with relevant PCR primers and conditions outlined above. Sequence identity of selected clones for both the mutated and wild-type products was confirmed by sequencing in a total volume of 10 μl using 1:8 dilution of BigDye® Terminator v3.1 Cycle Sequencing kit (Applied Biosystems, Warrington, UK) with vector primers directed against the SP6 or T7 promoter regions and following manufacturer’s instructions. Sequencing reactions were purified...
by ethanol precipitation and run on an ABI Prism 3700 DNA Analyser (Applied Biosystems, Warrington, UK). Transformants with the correct insert were placed into long term storage at −80 °C in 20% glycerol and plasmid generated from a single selected colony by large scale plasmid prep using Plasmid Midiprep System (Promega, Southampton, UK), following manufacturer’s instructions. Plasmid was quantified by absorption at 260 nm using a DU 640 spectrophotometer with a 50 μl cuvette (Beckman Coulter, High Wycombe, UK) to give a mass concentration and converted to copies using the following formula:

\[
\text{Number of copies of plasmid/μl} = \frac{\text{μg plasmid DNA calculated from 260 nm absorption/weight of plasmid + insert} \times \text{Avogadro’s number}}{\text{μ mole/molecular weight of plasmid}}
\]

After linearization with NotI (Promega, Southampton, UK), dilution series of both wild-type and mutant plasmid were used to generate standard curves and/or as reference spike for real-time experiments.

2.4. DNA extraction

DNA was extracted from soil samples by the method of Deng et al. (2010). Briefly, 1 g of soil was suspended in a solution of 0.12 M Na₂HPO₄ 1% SDS to form a slurry, 1 ml of which was subjected to bead beating. A reaction containing 10⁷ copies of the wild-type plasmid was added by inactivation by the addition of RQ1 stop solution and heating at 80 °C in 20% glycerol (Applied Biosystems, Warrington, UK) and the following conditions: an initial denaturing step at 95 °C for 15 min was followed by 40 cycles of 95 °C for 10 s, 54 °C for 10 s, 72 °C for 10 s and a single acquisition step at 81 °C for 5 s. This was followed by a melt curve: 95 °C for 10 s, 65 °C for 15 s then heating to 95 °C at a ramp rate of 0.1 °C/s, with continuous acquisition. A reaction containing 10⁵ copies of the wild-type plasmid was included in each PCR run as a calibrator to allow comparison to the wild-type calibration curve. For relative quantification, a second PCR was performed on each soil-DNA extract using the reference spike primers (Mut-F and Mut-R) under the above amplification conditions and using mutated plasmid as a calibrator with the reference calibration curve.

2.7. Calibration curve generation

Mutated and wild-type plasmids were diluted to produce standards of known copy-number to allow the generation of curves containing 10⁻¹⁻¹⁰ copies per reaction in a tenfold dilution series. Triplicate replicates of these standards (mutated and wild-type) were amplified, in the presence of DNA free soil extract to give an equivalent amount of co-extracted material, using the appropriate primers and the supplied software (Roche Diagnostic Systems, Burgess Hill, UK) to produce and store calibration curves for mutated and wild-type templates in a soil extract background.

2.8. Absolute and relative estimation

Absolute estimation using real-time PCR is based on direct estimation of copy number by comparison with a standard curve of diluted template. Relative estimation involves the separate amplification of the target gene and a reference of known copy number, here in the form of a ‘spike’ added at the DNA extraction stage. Both were assessed using the supplied software (Roche Diagnostic Systems, Burgess Hill, UK) applying the second derivative maximum method to quantify template DNA.

To estimate the number of 16S copies per gram of dry soil using absolute quantification it is necessary to take into account losses expected during DNA extraction. With the DNA extraction method applied here the following factors have to be taken into consideration: the actual dry-weight of soil extracted, the proportion of water in the soil extracted, losses involved in avoiding carry-over during the phenol:chloroform extraction and any dilution of the DNA during precipitation or post extraction. These factors were calculated to produce a multiplication factor specific for each sample. This factor was used to convert the absolute estimation of 16S copies into an estimation of 16S copies per gram dry weight of soil extracted. As it is impossible to measure the exact loss during the precipitation or the phenol:chloroform stage, we were forced to assume full recovery leading to an inevitable underestimation of copy number.

Relative quantification of the target DNA was assessed using the supplied software (Roche Diagnostic Systems, Burgess Hill, UK) for advanced relative quantification using default settings comparing it to stored standard curves for both the PCR reference and target. Relative Quantification compares the levels of two different target sequences in a single sample (16S and mutated 16S spike) and expresses the final result as a ratio of these targets. This ratio is used to calculate the number of gene copies per gram dry weight soil using (Spike copies × ratio)/dry weight soil added.

2.9. Statistical analysis

The data to be analysed comprised measurements of bacterial population size on samples of soil taken from locations with differing soil types. Four sub-samples were taken from each location’s material and, for each of these sub-samples, a measurement of copy number per gram d.w. calculated using both the absolute and relative real-time PCR techniques. These eight estimates from fourteen locations gave a total of 112 real-time PCR measurements. Four further sub-samples from each location were used to provide measurements of soil bacteria using ELFA. The mean of these four values at each location provided 14 values used as a gold standard against which comparisons were made. Both ELFA and real-time PCR data were transformed to log base 10 to
stabilise variance prior to analysis. Initially separate tests of the relationship of each real-time PCR measurement versus ELFA were performed using Analysis of Variance (ANOVA). Measurements from each technique were analysed with a term fitted for ELFA as a covariate and resolving the residual sum of squares into lack of fit and pure error (Draper and Smith, 1998). Deviations among the four copy number values for the same soil about their mean represented ‘pure error’ and were the result of variation within the same experimental material and a lack of repeatability of the measurement procedure. Deviations of the observed mean copy number for each soil sample from the line fitted to the covariate, ELFA, were referred to as ‘observed lack of fit’ and measured the goodness of fit of this line. This lack of fit variation gave a measure of validity of the technique in its agreement with the assumed ‘gold standard’ of the ELFA values. The estimated variances for pure error and observed lack of fit were then compared between the two real-time PCR approaches with an F-test to determine if there were any significant differences between the two molecular quantification approaches.

Since the measurements of the absolute and relative techniques were obtained from the same material in each of the four technical replicates for each soil sample, these did not provide independent estimates of the pure error variance. Similarly, there was no independence in the estimates of variance of the soil sample means about their respective fitted lines. This problem was addressed by using a mixed modelling approach on the combined datasets. In this approach pure error had the same interpretation as with the ANOVAs; however the observed lack of fit variation was decomposed into contributions due to the pure error and the further variation of soil sample means not attributable to pure error ‘underlying lack of fit’. Both pure error and underlying lack of fit were modelled as random effects but with covariance structures chosen in order to contain the relevant variance components and to address the lack of independence in the samples. In this way, we could formally test for differences in the variance components between the two techniques. Additional benefits of this approach were formal statistical tests for differences in the fixed effects of mean copy number recovered between the techniques and also differences in the slope of the fitted linear relationship to ELFA.

Three models were fitted. The full model had, within each random effect, separate variance components estimated for each technique as well as a covariance term. Two nested models could then be fitted, each testing one of the two hypotheses: that there is no difference between absolute and relative techniques in (i) the pure error variance components or (ii) the underlying lack of fit variance components. In each sub-model this was achieved by constraining the variance components for the two techniques to be equal. The difference in deviance could then be tested against that of the full model and its statistical significance determined against a chi-squared distribution with 1 degree of freedom since the number of parameters in the model for random variation differed by one.

The ANOVAs were fitted using GenStat for Windows, 14th Edition (VSN International Ltd., Hemel Hempstead, UK) as were the mixed models using the Residual Maximum Likelihood (REML) facility.

3. Results and discussion

3.1. ELFA

To assess the suitability of absolute and relative quantification of 16S copy number to estimate soil bacterial populations it was necessary to compare these molecular-based estimations against an established technique. The use of ELFA to estimate total bacterial populations in soil has previously been established (Schutter and Dick, 2000). ELFA analysis was used in preference to the more routinely employed phospholipid fatty analyses (PLFA) because ELFA analysis is more rapid and both PLFA and ELFA approaches are comparable in their ability to discriminate between microbial communities (Hinojosa et al., 2005; Griffiths et al., 2007). ELFA was performed on each soil type to obtain an estimation of total biomass based on lipid content and this data was used as a baseline to assess the performance and accuracy of real-time data. There was excellent reproducibility in lipid data over four replicates from each soil sample (Table 4), sufficient to justify the amalgamation of the replicates into a mean-value to investigate the variability within and between the two real-time quantification approaches and confirming the uniformity of the replicated soil samples. There was significant variation between soils in biomass estimation by ELFA (see Table 4), ranging from 10.20 to 623.29 nmol g⁻¹ dry weight soil (dws). Variation in biomass related to bacterial population sizes has been reported by others. For example, Hinojosa et al. (2005) in a single soil type observed similar variation when comparing polluted, reclaimed and pristine soils with both ELFA and PLFA methods with extracted ELFA between 22 and 184 nmol g⁻¹ dws with pristine soils exhibiting concentrations at the higher end of the range. There are few studies comparing across soil types; Frostegård and Båth (1996) analysed a range of soils from forest (Spruce and Beach), grassland, garden and arable fields and using marker PLFA representing the bacterial populations estimated biomass to be between 164 and 1739 nmol g⁻¹ organic matter in the different soils. Taking account of the measured % organic matter in each of these soils, these values convert to 34.4–516.1 nmol g⁻¹ dws. These estimates are within the range determined by this study but direct comparison is difficult due to their application of PLFA as opposed to ELFA techniques. Additionally, Frostegård and Båth (1996) could find no correlation either between the amounts of bacterial biomass as estimated by PLFA and pH or between the amounts of bacterial PLFA and the organic matter content of the soils studied similar to the patterns found here.

3.2. Real-time PCR

Calibration curves for both the target and spike were produced using a linearisled plasmid containing a PCR fragment of either E. coli 16S or a mutated version respectively. The PCR efficiency, in a DNA-free soil extract background, of both the wild-type and mutated PCR targets was high confirming the robustness of both PCRs. The calibration curves for the mutated template was linear over nine orders of magnitude (10⁻⁸–10⁹), the calibration curves for the wild-type template was linear over eight orders of magnitude (10⁻⁸–10⁻¹). This difference is likely due to absence of template from the PCR kit as the variation did not increase with decreasing template copy number. The sensitivity of the calibration curves reliably attained 10⁶ copies per reaction (Table 2).

3.3. Efficiency of recovery of DNA from soil as estimated by real-time PCR

A number of groups have used real-time or competitive PCR to investigate the efficiency of recovery of DNA from soil or sediment extractions (Table 3). These studies used a range of added targets to estimate the efficiency of recovery and estimation. Targets include genomic DNA, bacterial cultures containing a high copy number plasmid and bacterial cells seeded into sterilised soil samples, and vary in extraction methodologies

<table>
<thead>
<tr>
<th>Copies</th>
<th>Wild-type mean Cq</th>
<th>Mutated mean Cq</th>
</tr>
</thead>
<tbody>
<tr>
<td>10⁸</td>
<td>10.3 ± 0.04</td>
<td>9.6 ± 0.15</td>
</tr>
<tr>
<td>10⁷</td>
<td>13.7 ± 0.05</td>
<td>13.4 ± 0.06</td>
</tr>
<tr>
<td>10⁶</td>
<td>16.0 ± 0.01</td>
<td>17.3 ± 0.26</td>
</tr>
<tr>
<td>10⁵</td>
<td>20.2 ± 0.04</td>
<td>20.6 ± 0.17</td>
</tr>
<tr>
<td>10⁴</td>
<td>22.4 ± 0.16</td>
<td>24.3 ± 0.01</td>
</tr>
<tr>
<td>10³</td>
<td>25.9 ± 0.03</td>
<td>27.6 ± 0.01</td>
</tr>
<tr>
<td>10²</td>
<td>28.2 ± 0.05</td>
<td>30.9 ± 0.29</td>
</tr>
<tr>
<td>10¹</td>
<td>27.8 ± 0.09</td>
<td>35.0 ± 0.01</td>
</tr>
</tbody>
</table>
Table 3
Comparison of DNA extraction/detection efficiencies reported by previous groups and this study for different soil types. Details of soil types (this study) are given in Supplementary Table 1. Efficiency (this study) is calculated as average % of spike recovered over 4 replications and standard deviation.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Extraction method</th>
<th>Quantification method</th>
<th>Target</th>
<th>Soil type</th>
<th>Efficiency of recovery (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lee et al., 1996</td>
<td>Chemical</td>
<td>Competitive PCR</td>
<td>E. coli DNA</td>
<td>Sterilised pasture soil</td>
<td>3.68 ± 0.11</td>
</tr>
<tr>
<td>Rose et al., 2003</td>
<td>Competitive</td>
<td>Wizard® touchdown PCR</td>
<td>E. coli DNA</td>
<td>Silt loam</td>
<td>34 ± 7</td>
</tr>
<tr>
<td>Tsai et al., 1991</td>
<td>Chemical</td>
<td>Membrane hybridisation</td>
<td>E. coli DNA</td>
<td>Southern California</td>
<td>60 ± 5</td>
</tr>
<tr>
<td>Mummy and Findlay, 2004</td>
<td>UltraClean</td>
<td>FC-PCR</td>
<td>pBR322-λ</td>
<td>Lake sediment</td>
<td>14.9 ± 16.0</td>
</tr>
<tr>
<td></td>
<td>FastDNA SoilMaster</td>
<td></td>
<td></td>
<td></td>
<td>28.3 ± 10.5</td>
</tr>
<tr>
<td>This study</td>
<td>Chemical</td>
<td>Sybr green real-time</td>
<td>1 × 10^10 linear DNA</td>
<td>Soil 1</td>
<td>10.93 ± 2.87</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PCR</td>
<td></td>
<td>Soil 2</td>
<td>9.13 ± 5.29</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Soil 3</td>
<td>4.10 ± 1.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Soil 4</td>
<td>7.08 ± 6.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Soil 5</td>
<td>2.37 ± 1.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Soil 6</td>
<td>5.84 ± 2.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Soil 7</td>
<td>6.09 ± 2.44</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Soil 8</td>
<td>11.48 ± 5.40</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Soil 9</td>
<td>5.64 ± 4.39</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Soil 10</td>
<td>5.67 ± 2.93</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Soil 11</td>
<td>13.66 ± 4.44</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Soil 12</td>
<td>2.59 ± 1.21</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Soil 13</td>
<td>2.61 ± 2.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Soil 14</td>
<td>8.19 ± 3.51</td>
</tr>
</tbody>
</table>

42 T.J. Daniell et al. / Journal of Microbiological Methods 91 (2012) 38–44

with one study comparing several methods. The majority of these studies do not speculate on the likely source of variation in the estimation or cause of losses although the choice of extraction method (Mummy and Findlay, 2004) and/or the amount of added target (Rose et al., 2003) can clearly play a role. Soil, when compared to cultures, as a starting material also negatively affects both the recovery and variability of extraction and subsequent PCR inhibition both between and within soil types.

Our studies found recovery efficiencies of the spike of between 2.61% and 13.66%, dependent on, but with no clear link to, soil-type, which is of a similar range to recoveries observed in other studies (Table 3). Analysis of our spike-recovery data suggests that limitations in the efficiency of the soil extraction driven by differential effects of different soils play a part. Fig. 1 shows a significant variation in recovery of spike between soil types (p < 0.001) but suggests that variation was not driven by different perceived organic matter content between soil types. For example the two peat soils (13 and 14) vary significantly. A high degree of variation is found in the recovery of this internal standard for the wide range of soil-types included in this study as well as the high level of variation in recovery between replicates. There was also no correlation between efficiency of recovery and the amount of soil extracted (data not shown).

This difference in performance is likely due to variation in the distribution of inhibitors between aliquots or variation in the performance of the extraction between aliquots. The observed variability in recovery as estimated by an absolute real time PCR reinforces the improved performance of the relative tool discussed below since, unlike an unspiked soil, a fixed amount of template for the test PCR reaction was added to the extraction removing the possibility that variation was driven by natural variation in population size.

Where applying absolute quantification we attempted to account for all measurable losses in the DNA extraction process and yet this variation of DNA recovery efficiency demonstrates the limitations of simple absolute-quantification for estimation of DNA copies present in samples of different soil-types. For relative quantification this variation is countered by the internal standard used, as demonstrated by the significantly reduced variation between replicates. This reduced variation allows a far more precise estimation of population size.

3.4. Estimation of soil-bacterial population size

ELFA and real-time PCR data were converted to log base 10 to ensure residual variance remained constant prior to comparing the absolute (Fig. 2a) and relative (Fig. 2b) real-time quantifications against the ELFA data. The results of the separate ANOVAs for each technique are shown in Table 5 and demonstrate that there was evidence of significant differences between both lack of fit and pure error variation.

Also shown in Fig. 2a and b are the fitted lines from the REML analysis using the full model with separate variance components estimated for each of the respective estimation techniques. This analysis demonstrates the improved performance of the relative technique on a number of criteria. Total copy number per gram d.w. recovered were at least an order of magnitude higher for all soil types, ranging from a factor of 11.4 for soil type 6 to a factor of 23.0 for soil type 3 (intercept for the relative line was 8.61 compared to 7.70 for absolute). The slope of the fitted relationship for the relative technique was also significantly steeper (p < 0.001) – slope parameter was 0.91 for relative technique and 0.75 for absolute. This steeper slope demonstrates the greater ability to discriminate amongst soil types of the relative technique compared to absolute estimates. For example, the mean detected copy-number for samples of soil type 3 was greater than for soil type 1 when using relative estimation (which agrees with the gold standard) but this ranking is reversed when using the absolute approach. These differences are likely due to differences in extracted inhibitors between soil types. Since REML estimates variance components for (i) pure error, or, the repeatability of each technique and (ii) underlying lack of fit we fitted models which tested for the statistical significance of any difference in performance. The pure error variance component was tested for significant differences between the techniques by comparing deviance between models where pure error was identical (Model II in Table 6) and where different variances were fitted (Full Model, Table 6). This showed a highly significant difference (p < 0.001) with the best fit model estimating the absolute technique’s variance component at 0.173 compared to 0.011 for the relative; there was also a high correlation of 0.68 between the estimates from the two techniques performed on the same technical replicates. This evidence shows the relative

![Fig. 1. Efficiency of recovery of added template, showing the recovered copy-number (E±06) of the spike for each soil-type estimated by absolute quantification (1 ± E±09 added prior to extraction). Bars indicate ±1 standard error, letters indicate significant classes generated by post-hoc testing using a Fisher LSD test following ANOVA where soil types was significantly different (p < 0.001). Soil types as described in Supplementary Table 1.](image-url)
technique’s ability to achieve more consistent results on repeated analysis of the same soil samples.

In the comparison of sample mean squares for observed lack of fit in the ANOVAs, although the stratum variances were highly different (as shown in Table 5, F ratio was 5.33), this included a contribution from the pure error variance in the lower stratum. REML decomposed this variance and allowed the testing of the contribution from underlying lack of fit alone. This was done by testing a model with separate variance components for underlying lack of fit for the two techniques (the Full Model in Table 6) against one constrained to have the same variance component for both (Model I in Table 6). The Full Model gave variance components of 0.0766 for the absolute and 0.0197 for the relative techniques but there was no evidence for a difference when compared with Model I which fitted a single variance component of 0.0228. The reduction of ‘noise’ coupled with the more significant positive correlation between the ELFA data and the relative quantification estimation of population numbers also would allow accurate comparison of populations between soil types which would otherwise have significant error attached due to differences in efficiency of estimation. This type of analysis would not be reliable with the absolute quantification approach due to the variance between repeats of the same soil type. The lack of fit between the ELFA data and the absolute quantification makes this approach less appropriate when working across soil types. The relative quantification approach is far more suitable for this type of analysis.

Published estimates of bacterial population size in soil are variable, even when targeting the same gene. For example, estimates of gene copy number in agricultural soil based on 16S-targeted real-time PCR range from $6 \times 10^8$ to $2.2 \times 10^9$ copies per gram d.w. soil (Dandie et al., 2007; Okano et al., 2004). As described in Section 3.3, this range can be due to different soil-types supporting different population sizes of bacteria, effect of soil-type on efficiency of DNA extraction and/or PCR amplification or on limitations inherent in the measurement approach. The contrasting results between absolute and relative estimations observed in this study suggest that the application of relative methods may control for a number of these causes of error. Our estimations for soil-bacterial populations are $3.2 \times 10^9$ and $3.5 \times 10^9$ for absolute and relative quantification respectively (see Table 4), averaged over all soil types. Our estimation by absolute quantification is higher than that published previously, due, we believe, to our having made stringent attempts to account for all known losses during the DNA extraction (Section 2.8). Our estimation by relative quantification is above the published range, due to the relative quantification approach accounting for losses and inhibitory affects and therefore allowing a far more precise estimation of population size.

4. Conclusion

The application of relative quantification methods is standard in expression studies where housekeeping genes are used to control for differences in extraction efficiency or assay performance. Here we demonstrate that the addition of an artificial spike to soil extractions allows relative quantification methods to be applied in such a non-defined starting material by providing a fixed term of reference. This allows both a significant improvement in the estimation of gene copy number and a decrease in the variability in the estimation. The difference between relative and absolute methods of estimation suggests that the application of absolute measures, commonly applied in soil studies, can...
underestimate the gene copy count by factors ranging from 11.4 to 23.0. Additionally, the generation of an artificial spike negates concerns relating to erroneous detection of added spike as part of the soil community under analysis.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.mimet.2012.07.010.

Acknowledgements

This work was financially supported by the Scottish Government Rural and Environment Science and Analytical Services Division. We would like to thank both internal and external reviewers for comments which we have improved the manuscript.

References


Table 6

<table>
<thead>
<tr>
<th>Model</th>
<th>Soil sample</th>
<th>Technical rep</th>
<th>Deviance</th>
<th>LRS</th>
<th>df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>v\text{re}</td>
<td>v\text{co}</td>
<td>v\text{re}</td>
<td>v\text{co}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full</td>
<td>0.0766</td>
<td>0.0197</td>
<td>0.0189</td>
<td>0.173</td>
<td>0.011</td>
<td>0.029</td>
</tr>
<tr>
<td>I</td>
<td>0.0228</td>
<td>0.0228</td>
<td>0.0098</td>
<td>0.199</td>
<td>0.012</td>
<td>0.033</td>
</tr>
<tr>
<td>II</td>
<td>0.0973</td>
<td>0.0040</td>
<td>0.0194</td>
<td>0.090</td>
<td>0.090</td>
<td>0.028</td>
</tr>
</tbody>
</table>


