‘Ready Mixed’, Improved Nucleic acid amplification Assays for the Detection of *Escherichia coli* DNA and RNA

Short Communication/Note

Jonathan S. McQuillan* and Matthew W. Wilson

National Oceanography Centre, European Way, Southampton, SO14 3ZH, UK.

*Corresponding author
Email: jonmcq@noc.ac.uk
Telephone: +44(0)2380 592 715

Abstract (49 words)

The selective amplification of *E. coli* nucleic acid sequences could be used for the early warning of faecal contamination in environmental samples. Modified assays for *E. coli* DNA and RNA markers are presented with improved integrity and performance over existing methods, and demonstrated using ‘ready mixed’, preserved reagent mixtures.

Main Text (~1,500 words)

The presence of *Escherichia coli* (*E. coli*) in the environment is considered globally as probable evidence of faecal contamination, and therefore it is subject to statutory surveillance for various water and food safety assurance practices [1]. *E. coli* detection methods are generally culture based, and the provision of timely results is limited by the bacterial growth rate; typically they take in excess of 18 hours [2]. This increases public health risk, particularly during short-lived, stochastic contamination events. Nucleic acid amplification is a culture-independent technique, used to detect microorganisms by *in vitro* replication (amplification) of their DNA or RNA sequences. These methods can be employed to detect and enumerate *E. coli* with a better selectivity than cell culture [3], and in a few hours or less. The detection of *E. coli* nucleic acids could complement existing culture-based methods, particularly as rapid, 'early warning' risk indicators.

In this study, two 'improved' nucleic acid amplification assays are reported for the specific detection of *E. coli* DNA or RNA sequence markers. Each assay has been demonstrated using a convenient 'ready mixed' format whereby complete and dry-preserved reagent mixtures were prepared in advance, and then activated by simply rehydrating the mixtures with a water sample containing *E. coli* nucleic acids. In one assay, the detection of *E. coli* DNA was achieved using a quantitative real time PCR (qPCR) method targeting a fragment of the *ybbW* gene, based upon a highly selective primer set originally described by Walker *et al* [3]. A prior limitation of this method was the use of SYBR Green DNA binding dye for real-time fluorescence detection during qPCR. SYBR Green will bind to any DNA sequence which can lead to false-positive amplification from the presence of non-specific amplification products or primer duplex formation ('primer-dimers') [4]. Here, a *ybbW*-specific fluorometric hydrolysis probe was designed and tested in combination with the existing PCR primers to reduce background fluorescence signals and the likelihood of false positive amplification. A second assay targets *E. coli* mRNA by amplifying a fragment of the *clpB* gene transcript sequence. The same mRNA target has already been employed for the specific detection of *E. coli* in water samples using isothermal Nucleic Acid Sequence Based Amplification or NASBA [5-7]. The mRNA encodes a heat-shock response molecular chaperone protein and is induced by gently heating the bacteria prior to RNA isolation, enabling sub-single-cell sensitivity [7]. However, NASBA, which was recently compared to qPCR methods [3], displays significantly higher run-to-run

variability than PCR and is unreliable at low template concentration. The reagents are also comparatively expensive and not widely available [8]. Accordingly, here the method was adapted to work using reverse transcription qPCR for indirect mRNA amplification with hydrolysis probe-based real time fluorescence detection. The oligonucleotides used in this study were designed using Geneious R10 software (Biomatters Ltd, New Zealand) and synthesised by Integrated DNA Technologies (IDT Ltd, UK). Hydrolysis probes contained a covalently linked Fluorescein (FAM) at the 5' end, and an Iowa Black (IABk) quencher at the 3' end, and a secondary, internal ZEN quencher, 6 nucleotides from the 5' terminus. The sequences are shown in Table 1.

Reaction mixtures were prepared in sterile, nuclease-free PCR tubes (LightCycler 8-tube strips, Roche Molecular Systems Inc.). The qPCR reactions were prepared to contain 20mM Tris-HCl (pH8.3), 100mM KCl, 1.5mM $MgCl_2$, 2mM each dNTPs, 80nM each primer, 40nM hydrolysis probe, 0.2% (w/v) Sucrose, 0.2% (w/v) Trehalose and 2U of GoTaq G2 DNA polymerase (Promega, UK); the final volume was 100µL. The RT-qPCR reactions were prepared using the GoTaq Probe 1-Step RT-qPCR System (Promega, UK) following the manufacturers recommended protocol except that the final 20µL reactions were mixed with 80µL of RT-PCR grade water to a final volume of 100µL. The mixtures were frozen at -80^{o}C, and then lyophilised using a sterile vacuum chamber in tandem with an a ZL(8L) Lyophilisation Instrument (SPScientific, UK). The vacuum chamber was decontaminated before use by applying DNAaway (Sigma, UK) and RNaseZap (Thermofisher, UK) chemicals according to the manufacturers recommended method, to remove contaminating DNA and RNase respectively. Lyophilisation took place for precisely 16 hours at <200 µBar, at which point the mixtures had formed a crystalline, white powder. Air for re-pressurisation passed through a 0.2 micron Sterivex filtration unit (Millipore, USA). Decontamination was assessed by no template control reactions showing null amplification. After varying periods of time in storage at room temperature, the lyophilised reaction mixtures were activated by adding a 20µL solution of *E. coli* DNA or RNA containing a known quantity of the target sequence. The DNA template was prepared from genomic DNA isolated from a type strain of *E. coli* (NCTC 9001), exactly as described by Walker *et al* [3]. RNA template was prepared by *in vitro*, T7 RNA Polymerase-driven expression of a truncated *clpB* gene sequence fused to a 5' T7 promoter sequence. Full details of the template preparation are given in the supporting information. The template samples were quantified, and stock solutions were prepared to contain an

estimated 10^5 to 10 copies per 20 microliters. Each qPCR reaction was carried out using a LightCycler 96 qPCR instrument (Roche, UK) with an initial denaturation of 95°C for 2 minutes, followed by 40 cycles of 95°C for 20 seconds and 60°C for 60 seconds. The RT-qPCR reactions were carried out as described above, except for the inclusion of an initial reverse transcription step of 42°C for 15 minutes prior to thermal cycling.

When prepared and performed exactly as described above, the preserved mixtures could be used to amplify between 10 and 10^5 estimated target sequence copies, as shown in Figure 1. Furthermore, the integrity and reliability of each method was improved when compared to prior versions. For example, the use of a *ybbW*-specific hydrolysis probe (this study) in place of SYBR green [3] for real-time detection of the *ybbW* target eliminated background fluorescence and false-positive amplification over a 40 Cycle PCR, as shown in the supporting information Figure S1. In addition, the detection of *clpB* mRNA by RT-qPCR (this study) in place of NASBA (multiple prior studies; [3, 5, 7]) markedly reduced the variability between replicate reactions and different runs, as shown in the supporting information Figure S2. Each 'improved' assay was able to detect at least 10 estimated copies of the target sequence, but when the template was diluted further only a portion of the replicate reactions generated amplification curves (data not shown), and so the limit of quantification (LOQ) was taken to be ≥10 estimated copies for each assay. The relationship between template sequence copy number and C_t was linear over 5 orders of magnitude, with a typical linear fit (R^2) of 0.9934 (*clpB* RT-qPCR) and 0.9991 (*ybbw* qPCR). The efficiency of the primers was determined using the method of Pfaffl [9] and found to be 1.94 (*ybbW*) and 1.83 (*clpB*), when using freshly prepared mixtures. Storage for up to 4 weeks did not impact the LOQ of the qPCR, albeit the amplification rate was reduced (C_t was increased). However, preservation increased the LOQ for the RT-qPCR to 100 estimated copies. After 4 weeks the amplification efficiency for each oligonucleotide set was 1.78 (*ybbW*) and 1.67 (*clpB*); the linear relationship (R^2 ≥ 0.99) was unaffected by storage. After 6 weeks in storage there was a significant loss in reagent activity, and only the samples containing the highest tested number of target sequence copies (10^5 per reaction) could be amplified (not shown). Accordingly, the 'shelf-life' of the preserved mixtures used in this study was considered to be up to 4 weeks.

The inclusivity and specificity of each oligonucleotide set (primers and probe) was evaluated to determine whether the modifications made to existing methods, as described in this study, had any impact on their selectivity for *E. coli*. This was done according ot the method of Walker *et al* [3] by PCR amplification of DNA sequences extracted from a panel of *E. coli* strains, and non-*E. coli* bacterial species, full details of which can be found in the supporting information Table S1. PCRs were carried out as described above, but using freshly prepared (i.e., not preserved) reaction mixtures. Each assay was 100% inclusive of 76 unique *E. coli* strains including 72 strains belonging to the ECOR collection, which is considered to represent genotypic variation in *E. coli* [10, 11]. The *ybbW* qPCR was able to exclude 22 non-*E. coli* bacterial strains. The *clpB* assay was mostly specific for *E. coli*, except that it was able to detect DNA extracted from closely related species *Shigella* spp., *Escherichia albertii* and *Escherichia fergosonii*. Overall, our results were in agreement with those reported by Walker *et al* [3], who carried out the same tests on unmodified versions of these assays, indicating that the changes described in this study did not negatively impact the selectivity of each method.

In summary, the methods described here can be used to amplify (detect) *E. coli* DNA and RNA sequence markers at concentrations ranging from ≤10 to 100,000 copies, with a strong linear correlation for quantification. Each method was based on existing, state of the art nucleic acid amplification tests for *E. coli*, but including critical modifications to improve integrity and reliability. The use of the dry-preserved reaction format constitutes a streamlined, one-step testing process, suited to automation, and where the potential for human error and contamination are significantly reduced. This work was funded by the Natural Environment Research Council, grant NE/R013721/1.

References

1. Edberg, S.C., et al., *Escherichia coli: the best biological drinking water indicator for public health protection.* Symp Ser Soc Appl Microbiol, 2000(29): p. 106S-116S.
2. SCA, *The microbiology of recreational and environmental waters (2014) - Part 3-Methods for the isolation and enumeration of Escherichia coli (including E. coli O157:H7).* Methods Exam. Waters Assoc. Mater., 2016.
3. Walker, D.I., et al., *A highly specific Escherichia coli qPCR and its comparison with existing methods for environmental waters.* Water Res, 2017. **126**: p. 101-110.
4. Zipper, H., et al., *Investigations on DNA intercalation and surface binding by SYBR Green I, its structure determination and methodological implications.* Nucleic Acids Res, 2004. **32**(12): p. e103.
5. Min, J. and A.J. Baeumner, *Highly sensitive and specific detection of viable Escherichia coli in drinking water.* Anal Biochem, 2002. **303**(2): p. 186-93.
6. Malek, L., R. Sooknanan, and J. Compton, *Nucleic acid sequence-based amplification (NASBA).* Methods Mol Biol, 1994. **28**: p. 253-60.
7. Heijnen, L. and G. Medema, *Method for rapid detection of viable Escherichia coli in water using real-time NASBA.* Water Res, 2009. **43**(12): p. 3124-32.
8. Honsvall, B.K. and L.J. Robertson, *From research lab to standard environmental analysis tool: Will NASBA make the leap?* Water Res, 2017. **109**: p. 389-397.
9. Pfaffl, M.W., *A new mathematical model for relative quantification in real-time RT-PCR.* Nucleic Acids Res, 2001. **29**(9): p. e45.
10. Ochman, H. and R.K. Selander, *Standard reference strains of Escherichia coli from natural populations.* J Bacteriol, 1984. **157**(2): p. 690-3.
11. Patel, I.R., et al., *Draft Genome Sequences of the Escherichia coli Reference (ECOR) Collection.* Microbiol Resour Announc, 2018. **7**(14).

Table 1. Oligonucleotides used in this study.

Name	Sequence (5' – 3')	Reference
ybbW forward primer	TGATTGGCAAAATCTGGCCG	
ybbW reverse primer	GAAATCGCCCAAATCGCCAT	
ybbW hydrolysis probe	FAM-CCGCCG[ZEN]AAAACGATATAGATGCACGG- IABkFQ*	This study
clpB forward primer	GCGACAATCCGGTCTTCA	This study
clpB reverse primer	AAATCCACATTTCTGACGAGG	[7]
clpB hydrolysis probe	FAM-CTTCCA[ZEN]GGCGAATCACTTTACCCGG-IABkFQ*	This study

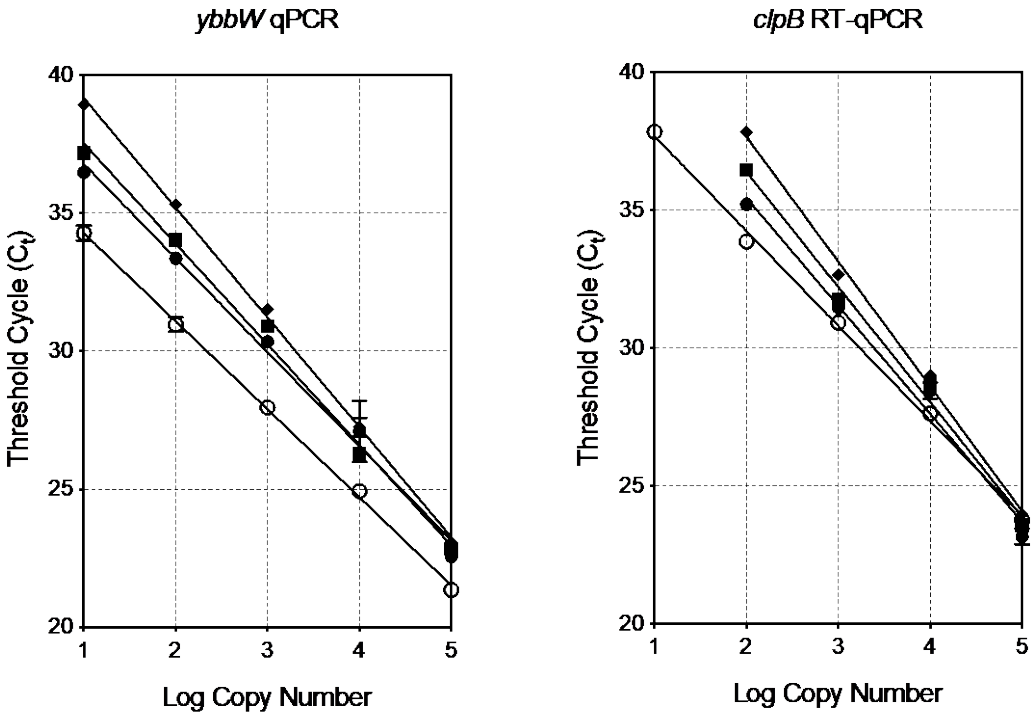


Figure 1. The oligonucleotides shown in Table 1 were used to amplify *ybbW* DNA sequences by qPCR and *clpB* RNA sequences by RT-qPCR using dry-preserved reaction mixtures which had been stored for up to 4 weeks without refrigeration. After 1 week (●), 2 weeks (■) or 4 weeks (◆) the mixtures were rehydrated with water containing DNA or RNA template at an estimated concentration of between 10 and 100,000 copies. For comparison, the open circles (**O**) indicate reactions prepared using fresh reagents which had not been preserved, but did contain an equivalent amount of Trehalose and Sucrose sugars. The results show the mean threshold cycle (C_t) versus template copy number from quadruplicate reactions. The error bars, where visible, show the standard error of the mean (n=4). No symbol represents a null amplification.

Supporting Information

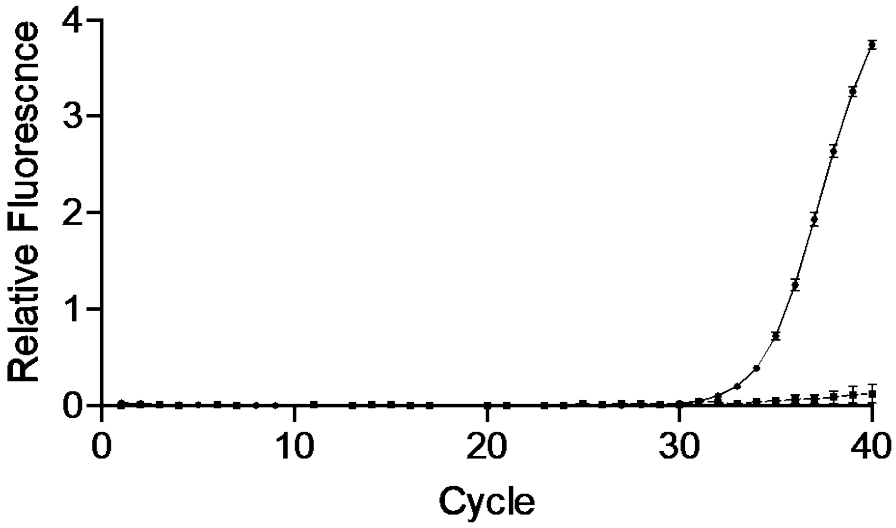
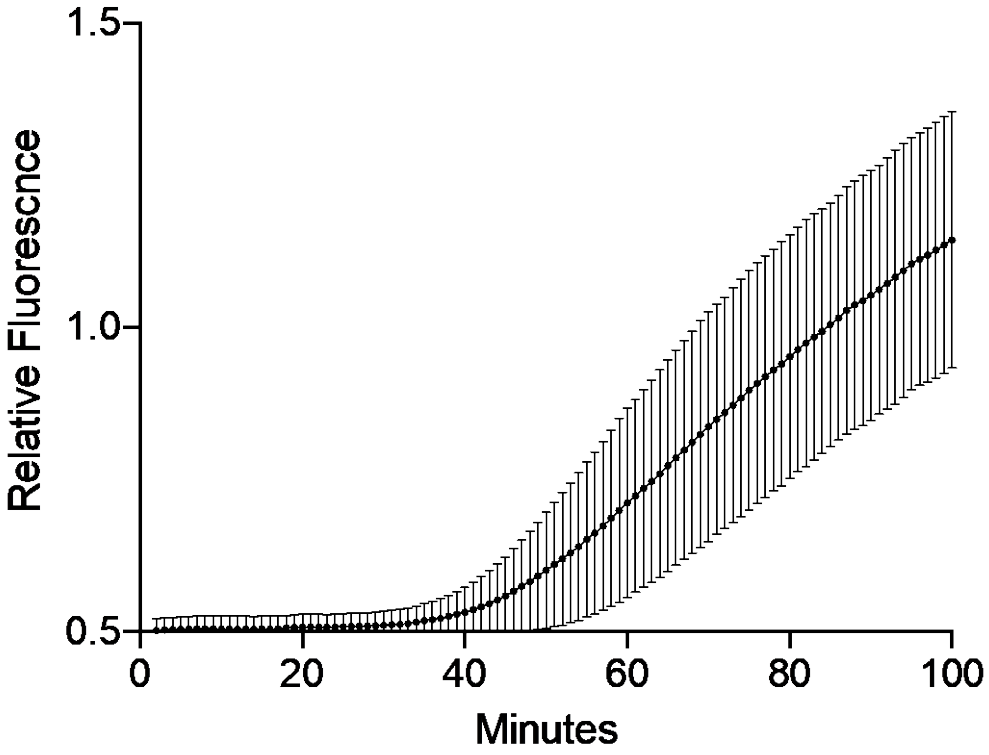


Figure S1. Fluorescence plots from qPCR reactions using the *ybbW*-specific primers and either SYBR Green (●) or the hydrolysis probe (■) for real-time detection. The same DNA template was added to each reaction, and was negative for the *ybbW* target sequence. In each case, use of agarose gel electrophoresis (not shown) confirmed that no reaction product had been formed. However, the use of SYBR green led to a false-positive result, probably due to the accumulation of primer duplex or non-specific amplification products during the later stages of the reaction. In contrast, use of the hydrolysis probe generated a correct (null) result over the 40 cycle reaction. The results show the mean of triplicate reactions. The error bars, where visible, show the standard error of the mean.



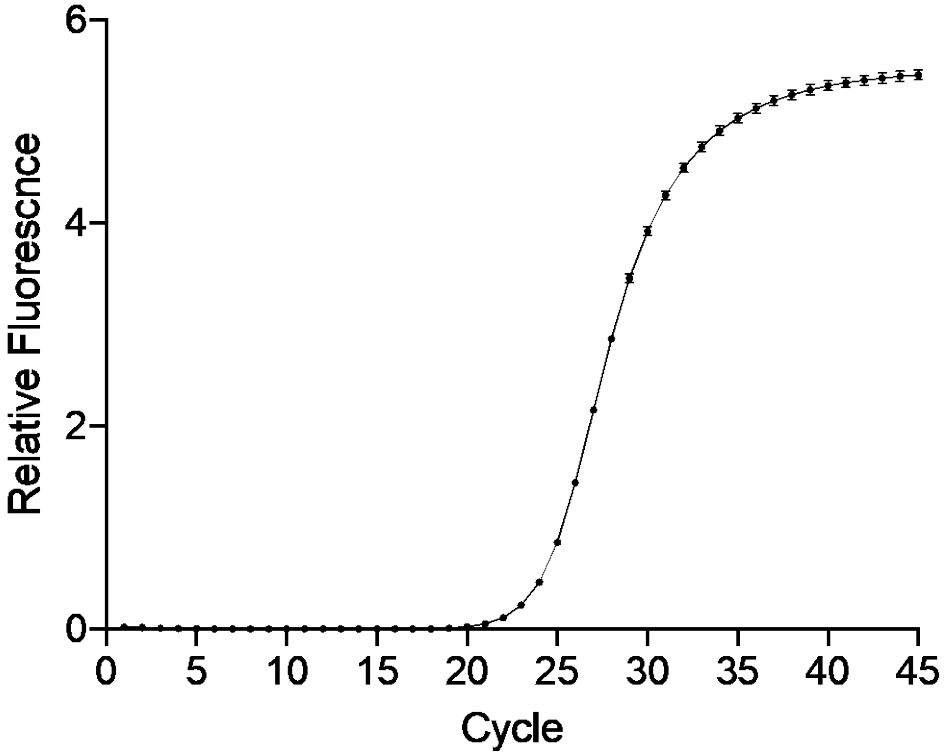


Figure S2. Amplification of *clpB* mRNA using NASBA (top image) or RT-qPCR (bottom image). The results show the mean from replicate reactions. The error bars show the standard deviation, n=5.

Table S1. Selectivity Testing

Species	**Collection**	**Strain**	***clpB* RT-qPCR**	***ybbW* qPCR**
Escherichia coli	ECOR	1-72	+ (72)	+ (72)
Escherichia coli	NCTC	9001	+	+
Escherichia coli	NCTC	12241	+	+
Escherichia coli	NCTC	13216	+	+
Escherichia coli	NCTC	12900	+	+
Aeromonas caviae	NCTC	10852	-	-
Citrobacter freundii	NCTC	9750	-	-
Citrobacter koseri	NCTC	10786	-	-
Enterobacter aerogenes	NCTC	10006	-	-
Enterococcus faecalis	NCTC	775	-	-
Enterococcus faecium	NCTC	7171	-	-
Escherichia albertii	NCTC	17582	+	-
Escherichia fergusonii	NCTC	12128	+	-
Escherichia hermanii	NCTC	12129	-	-
Escherichia vulneris	NCTC	12130	-	-
Klebsiella pneumoniae	NCTC	9633	-	-
Listeria monocytogenes	NCTC	11994	-	-
Pantoea agglomerans	NCTC	9381	-	-
Pseudomonas aeruginosa	NCTC	10332	-	-
Salmonella bongori	DSMZ	13772	-	-
Salmonella Nottingham	NCTC	7832	-	-
Shigella boydii	DSMZ	7532	+	-
Shigella flexneri	DSMZ	4782	+	-
Shigella sonnei	DSMZ	5570	+	-
Shimwellia blattae	NCTC	12127	-	-
Vibrio cholera	NCTC	8042	-	-
Vibrio parahaemolyticus	NCTC	10885	-	-

(+) positive amplification; (-) negative amplification.

DNA and RNA Template Preparation

Genomic DNA standards were prepared from an *E. coli* type strain (NCTC 9001), according to the method of Walker *et al*. Briefly, the *E. coli* were revived from storage at -80^{o}C and cultured in Luria Broth at 37^{o}C. An exponentially dividing culture was harvested by centrifugation (5,000 g for 5 minutes) and resuspended in Maximum Recovery Diluent for 1 hour. Then, the cells were centrifuged again, and the cell pellet was used to prepare a genomic DNA extract using the GeneElute Bacterial Genomic DNA Isolation Kit (Sigma, UK). The DNA was eluted and stored in Tris-EDTA buffer (pH 8.0) at -20^{o}C. The mass of DNA in the extract was estimated by Qubit Fluorometric Quantification in tandem with the Qubit dsDNA high-sensitivity assay kit (ThermoFisher, UK), and used to estimate the number of genome copies. A series of genome copy number standards were prepared by diluting the extract in RT-PCR grade water (Promega, UK) to between 10 and 10^5 copies per microlitre. Standards were prepared from a stock DNA solution immediately prior to use.

RNA Standards were prepared as follows. First, a fragment of the *clpB* gene sequence was amplified by PCR using forward primer 5'-**AATTCTAATACGACTCACTATAGGG**AGAAGGTACTGGACGGCGACAATC-3' and reverse primer 5'-ATGGAGAAACACTCGGTGTC-3'. The forward primer had a T7 RNA Polymerase promoter sequence (shown in bold), followed by a short 'spacer' (shown underlined), upstream of the target-binding sequence, which generated a dsDNA PCR product which could be used to synthesise template RNA using T7 RNA Polymerase. The PCR product was purified using the GeneElute PCR Purification Kit (Sigma, UK) and used directly for RNA synthesis using the Hi Scribe T7 RNA Synthesis Kit (NEB, USA) according to the manufacturers recommended protocol. The synthesised RNA was purified using the RNeasy Mini Kit (Qiagen, UK), and contaminating dsDNA was eliminated using RQ1 RNase-Free DNase (Promega, UK) according to the manufacturers recommended protocol. DNA elimination was confirmed by a null Taq-based PCR. The DNA-free RNA was subsequently purified a second time using the RNeasy Clean-up procedure. The mass of RNA was estimated using a BioAnalyser Instrument, and the RNA 6000 Nano Kit (Agilent, UK). This was used to estimate the number of copies based upon the RNA sequence. A series of RNA copy number standards were prepared by diluting the RNA sample in RT-PCR grade water to between 10 and 10^5 copies per microlitre. Standards were prepared from stock RNA solution immediately prior to use.