Chapter 32 Quantification of host and phage mRNA expression during infection using
real-time PCR

Dr Martha R. J. Clokie,
Department of Infection, Immunity and Inflammation,
Medical Sciences Building,
University of Leicester,
Leicester
LE1 9HN, U. K.

+44 (0)116 252 2959 (phone)

+44 (0)116 2525030 (fax)

E.mail : mrjc1@le.ac.uk

Abstract

Real-time, or quantitative PCR is a valuable technique useful in bacteriophage research to quantify the abundance of phage or host gene transcripts. It can be used during the infection cycle both to monitor the expression of individual viral transcripts and to compare relative gene expression levels throughout the infection cycle. It is fairly economical to conduct and is useful in bacteria-phage systems where obtaining high yields of RNA is problematic. To perform real-time PCR it is simply necessary to know the DNA sequence of the genes to be monitored, to have accurately quantified mRNA and good quality cDNA, and access to a light-cycler. Although this chapter briefly reviews the basic principles of real-time PCR, the emphasis is on aspects of technique that are specific to the study of bacteriophage transcriptomics. These include (1) the

selection of the target gene, (2) the choice of calibrator and reference genes, and (3) RNA isolation for cDNA synthesis and (4) subsequent analysis of samples. This chapter should also be useful to those wishing to amplify genes from other types of templates such as metagenomic DNA or RNA extracted either from filtered samples or from agarose gels.

Keywords

Real-time PCR, SYBR green, probes and primers, endogenous control, calibrator

1. Introduction

Real-time or quantitative PCR combines PCR (chapter 27) with spectrophotometry such that an increase in DNA concentration is measured in 'real-time' as opposed to standard PCR, where, the product size and approximate quantity is measured on an agarose gel at the end of the reaction. Unlike conventional PCR, the conditions of the reactions are standardised for all transcripts to allow multiple transcripts to be examined during each analysis. Real-time PCR works best when the region amplified (the amplicon) is small. In addition to the usual components for PCR, a fluorophore (either SYBR green or a fluorescent probe) is included in the reaction mix and the reactions are run in clear plastic so the amount of fluorescence can be measured during the relevant stage in the amplification procedure (after polymerization has been completed). The fluorescence level increases exponentially with the concentration of PCR product and is measured and plotted logarithmically against the cycle number. A threshold value (C_T) is then set so that the samples can be compared. This threshold value can be at any point in the exponential phase of the amplification (Figure 1). To give an example of how C_T values allow a comparison of transcript abundance, at any point in the exponential phase, if a particular transcript has a C_T value of 15 at a particular threshold, then half as much template at that threshold would give a C_T value of 16, and half as much again would give a C_T value of 17 and so on.

[Insert Figure 1 near here]

Depending on the experimental design and plate layout, analysis of these data allows either the absolute or relative amount of a gene transcript in a sample to be estimated. In absolute quantification, the absolute quantity of a target sequence from a single nucleic acid from a particular sample is determined whereas in relative quantification the amount of target sequence is determined relative to the amount of the same (or different) target sequence in a calibrator sample. In bacteriophage transcriptomic research, this calibrator is likely to be a particular point in a time course. Thus for example we may wish to determine how much transcript encoding the major capsid protein is present ten minutes after infection compared to one minute after infection. While absolute quantification is useful in some scenarios, relative quantification is generally more useful in bacteriophage transcriptomics as it is often necessary to know how much of a transcript is present relative to a particular time point. Relative quantification has the advantage that accurate comparisons between samples can be determined over several plates without using standard curves on each plate.

As with real-time PCR in any system, the major decision to be made is between the technology based on SYBR Green or that which uses specific probes and primers (1). SYBR Green binds to double-stranded DNA where it fluoresces when it is bound. There are many different designs of specific primers and probes, made by an equally large number of biotechnology companies (e.g. Applied Biosystems and Sigma). In general the makers of specific real time PCR machines recommend their own particular chemistries, (though they are often not the cheapest), that should be compliant with most machines. The probes used in real-time PCR anneal to the DNA in the region between the

two primers, and in general they have two dyes, one which fluoresces and one which is a quencher molecule. Fluorescence only occurs, and thus is detected, when the polymerase cleaves the fluorescent dye from the probe (1). There are many variations on this theme.

The advantage of SYBR green is that it is much cheaper than using probes and primers and therefore numerous genes can be examined simply for the extra cost of additional primer pairs. It is also useful if the gene is variable in the region between the primers, as only the DNA which encodes the primer sequences that needs to be conserved. This is in comparison to using probes as well as primers, as the probe hybridises to the region between the primers so also has to be conserved. Thus SYBR green could be used to quantify the amount of template present for a mixed population a gene where some versions have a mutation between the primers. The major disadvantage to SYBR green however is the lack of specificity, the greater risk of contamination and the longer time-period necessary to optimize the experiments. However, the specificity can be analysed using a meltcurve (T_M) analysis following amplification. Here the finished plates containing the PCR products and SYBR green are heated to 95 °C. During this time the instrument records the drop in SYBR green fluorescence that results from the dissociation of double stranded DNA, primer dimers will have a lower melting temperature than that of the product, and more than one curve would represent multiple products.

The main advantage to a probe based approach is its specificity, as the fluorescent signal is generated from the binding of the probe and target rather than simply to all double-

stranded DNA. Less optimisation is therefore required for the reaction and generally contamination is much less of a problem. A further advantage to probes and primers is that duplex experiments can be performed, where two products are amplified using different fluorescent probes (see note 1). A disadvantage is the cost, however when the cost of a researchers time is taken into account this may be negated.

The widespread infection of all bacteria with bacteriophages is increasingly apparent. It is often desirable to establish at what stage bacteriophages are in during their infection cycles. This important information can be obtained from an examination of the infected cell transcriptome. Real-time PCR is a highly sensitive tool with which to probe this in order to monitor the presence of transcripts either throughout a lytic infection cycle, or during infection by a lysogen. It can be used independently to monitor transcript expression, or as a tool to corroborate and calibrate a micro-array data set (2, 3).

Although bacteriophages must inject their DNA, replicate their genomes and construct protein coats, the order and timing of these events has only been established for a few phages [e.g. T4 (2, 3)]. Furthermore, expression profiles of the plethora of host and genes of unknown function which are found inside newly sequenced genomes is generally uncharted.

Real-time PCR is a standard technique which is increasingly used for quantification purposes. There is a large body of literature detailing developments on chemistries, experimental design and analysis. This was nicely summarized in a recent review by

Bustin (4). Useful introductions to the principles and applications of real-time PCR can be found in the literature provided by manufacturers of Real-time PCR machines, (e.g. from Applied Biosystems, Corbett, and Stratagene). There is also a recent excellent and comprehensive book on the subject (5). However, relatively little has been written specifically on the use of real-time PCR in bacteriophage-host relationships and this chapter represents the first methods overview on the subject. Real time PCR can also be applied to examine the quantity of phage DNA from a range of templates. In which case the protocol is even easier; simply follow the guidelines in Chapters 23 and 26 to prepare a high quality DNA template, then all else applies. Other templates that are suitable for real-time PCR analysis include metagenomic samples (where for example the amount of a particular phage, or phage gene in a particular viral or bacterial DNA fraction could be determined (6, 7).

2. Materials

2.1 Good quality template

The extraction of mRNA is covered in chapter 32. Although real-time PCR can detect DNA concentrations as low as 1 fg, for routine real-time PCR best results may be obtained when using between 1 – 100 ng of mRNA to be transcribed into cDNA. For each reaction well on the RT-PCR plate between 1-10 ng of cDNA will be used.

2.3 Primers and probes

Primers can be ordered from your regular supplier (e.g. Invitrogen), and probe and primer combinations can be obtained from several biotech companies (e.g. Applied Biosystems).

A variety of online resources can help you to design appropriate primers (see <http://molbiol-tools.ca/PCR.htm>). Most companies offer two routes to obtaining probe and primer assays with either the option of designing primers oneself, or, they will design the assays for you. Often only the latter is guaranteed to work, but you have the option of selecting the approximate gene region in which you would like the probe to be placed, as it is possible to mask all areas of the gene that you do not wish primers/probes to hybridize. The products must be between 60 – 200 bp and the primers must anneal at 60 °C (8). There are various software programmes which specifically design primers for real-time PCR (e.g. Primer Express from Applied Biosystems).

2.4 Master mix. This contains the reagents that are present in standard PCR, including polymerase, buffer, $MgCl_2$, dNTPs and water and SYBR green (when used). This mix generally is at 2X strength and can be ordered from several biotech companies (e.g. Applied Biosystems, Qiagen, Sigma) (see note 2).

2.5 Nuclease-free water (e.g. Sigma)

2.6 Plates and adhesive covers or tubes. Either 96 or 386 well plates, if access to a robot is available. Adhesive covers must be heatproof and of optical quality. Some machines (e.g. Corbett) use individual tubes, and may take up to 72 individual tubes per reaction.

2.7 Light cycling PCR machine

Several companies now make light-cyclers, that are based on a Peltier block in which a 96 well plate sits. These include Applied Biosystems, Qiagen, Eppendorf. The alternative is to have a rotary chamber in which the tubes sit and air is pumped in and out at the correct temperature (e.g. Corbett). The data acquisition is particularly fast in the latter systems and the uniformity of heat distribution is greater.

2.8 Bench top vortexer

2.9 Laminar flow hood and related precautions

It is important to minimize exposure of samples to nucleases and any form of contaminants. To avoid any problems from the outset, it is highly recommended that all real-time PCRs are setup in a laminar flow hood which has been sprayed with anti-nucleases. Gloves must be worn at all times, all plastic ware must be nuclease free and pipette tips should have filter plugs.

3. Methods

3.1 Time point selection and phage infection

It is important to make sure that the time points chosen for a real-time PCR are representative of the different growth phases of the phage in question. To probe the transcriptome, one should have prior knowledge of the absorption rate of the phage. This will allow the infections to be synchronised and prevent the signal being muddled by the noise from second rounds of infection (Chapter 16). Other physical parameters should be measured such as the eclipse and latent periods (Chapter 16). Time points can therefore be taken at different phases of the cycle which will range from minutes (i.e. for T4 infecting *E. coli* to hours or even days for example in S-PM2 infecting *Synechococcus*). Procedures for phage infection and RNA isolation over a time course generally should follow the considerations described in Chapter X. In particular the RNA must be preserved rapidly and stabilised at each time point where reagents such as RNAlater can be extremely useful.

3.2 Target gene selection

To undertake real-time PCR, the sequence of the transcript of interest must be known. The transcript may either be part of a gene with known function, or possess homology to a gene with a known function. Alternatively its function may be unknown, and therefore its expression profile may at least show at what point during the infection cycle it is transcribed.

If the researcher is studying a new phage-bacteria system, then it may be useful to obtain information relating to genes where the expression profile is known in related phages. This may include structural genes or known ‘early’ or late genes. These data can give a framework of expression in which to compare the expression of novel genes. It may be that the researcher has a specific gene in whose function and expression profile they have an interest. However it may be that the question of interest is more general. For example, the researcher may simply be interested in when phage capsid proteins are synthesized, in which case it would be sensible to monitor the expression of the major capsid protein. For T4, the expression profiles of most genes has been determined (2, 9) and proves a useful starting point when evaluating gene expression of under-studied T4 type phages infecting other bacteria.

3.3 Control calibrator gene selection

One of the main difficulties with bacteriophage transcriptomics is that there is not necessarily one gene that is constitutively expressed throughout the infection cycle. For example, T4 switches off the replicative machinery of its host shortly after infection (9). Being reliant on its host as an energy source, the cyanobacterial phage S-PM2 appears not to have this method of total host switch off (10, 11). As is the situation in other transcriptional studies, the ribosomal machinery appears to represent the best of an imperfect solution. Certainly it seems that in the case of S-PM2 the amount of ribosomal RNA as represented by the 16 S transcript remained constant at all time points. One problem with using the ribosomal mRNA as an endogenous control (calibrator) is that

because it is present at a much greater abundance than any of the bacteriophage or host genes of interest it needs to be analyzed using a different threshold sensitivity compared to all of the other genes. Other control genes commonly used include *rpoD*, *gyrA* and others. None is perfect and each may present specific concerns depending on the phage-host system under study. Having available the sequence for a candidate calibrator gene from the bacterial host will often determine what is selected. The highly conserved regions of 16S rRNA typically circumvent this problem. When designing primers for the calibrator, one should follow the same considerations used for the target gene primers.

One way to circumvent the problem of identifying a gene with which to normalize data is to normalize to the total amount of RNA extracted from the cells. Unfortunately, this approach also has its pitfalls as it is difficult to accurately quantify the cDNA, and discrepancies may arise between samples collected at different time points and unequal conversion of mRNA to cDNA. Although these are valid concerns, I recently analyzed a data set consisting of the expression of 12 bacteriophage genes and 5 host genes during an infection and obtained very similar results regardless of whether I normalized to 16S or to total mRNA (unpublished). Thus although normalizing to either 16S mRNA or to total mRNA is not perfect, it is reassuring they essentially report the same information.

The choice of calibrator is always difficult, and not a problem specific to phage-bacterial work, the researcher should probably start by using 16S as a calibrator, and if possible test the levels of expression of host genes that they believe may not be influenced by phage infection.

3.4 Primer or primer and probe optimisation

It is important to test all primers or probes and primer combinations before analyzing the valuable cDNA samples obtained from the experiments. One must ensure that amplification is equally efficient and reports correctly across a range of template concentrations. If the researcher is using relative quantification, then after the standard curves have been carried out once, it is not necessary to repeat them on each plate.

For each gene to be assayed, make a serial dilution of either cDNA or phage genomic DNA, so samples contain 10, 5, 2.5, 1.25, 0.625 and 0.3125 ng of template. Add the appropriate amount of probe/primer and master mix. Run the reaction and check that the slopes of the amplification curves are parallel in log view and that for each sample that reduced by half in the quantity of cDNA there is an increase of 1 in the C_T value.

If the complete genome sequence of the bacteriophage is known, and if it has one copy of all the genes for which you wish to monitor the expression, then it serves as a natural template with which to compare the efficiency of amplification of gene transcripts, and to establish a standard curve. If the genomic template is amplified equally efficiently for each gene primer set, then the threshold value should be identical for each gene. One then assumes that equal efficiency will occur with the cDNA as template, although due to differences in transcript abundance, this can not always be confirmed.

3.5 Plate design

Having established that the primers and probes work and report evenly for all target and calibrator genes, one can then proceed to the experiment. Table 1 shows an example of a typical setup for the relative quantification of a number of bacteriophage genes using real-time PCR. All assays are performed in triplicate and should contain a negative control for each transcript. The negative controls should contain everything but template. If using relative quantification then fluorescent controls should also be present to allow cross plate comparisons. Thus the setup on this plate would allow a regular relative quantification, cross plate comparisons, and an estimation of the actual number of bacteriophage transcripts (see below).

[Insert Table 1 near here]

The experiment shown in Table 1 had five-time points; 0, 1, 3, 6 and 9 hours after infection. This covered the latent period for the phage S-PM2 infecting *Synechococcus* WH7803 at an MOI of 10. The expression profiles were monitored for one bacteriophage gene of interest (D1p) and three host genes (D1h, D1I1 and D1I2) and the ribosomal RNA (16S) was used as an endogenous control.

The plate also dedicates 12 wells to fluorescent controls. These controls consist of four different bacteriophage genes which previously showed equal amplification from purified S-PM2 DNA. Their presence here is to allow easy cross plate comparisons as 10 ng of

purified bacteriophage DNA from the same stock was used for each fluorescent control. As these measurements do not vary, the threshold at which the C_T value is chosen can be maintained throughout all plate comparisons. When further host and phage genes are examined from these samples the C_T values will be comparative to each other, by checking the C_T values of these fluorescent genomic DNA controls.

These fluorescent controls have a second bacteriophage-specific purpose as calibrators. Because it is possible to calculate how many phage genomes must be present in a known concentration of DNA (see note 3), the number of transcripts present in the expression work can be calculated by comparison to the genomic viral DNA samples.

One final point regarding fluorescent controls is that the use of an assay (e.g. primers and probes for gene 47) as a fluorescent control on gDNA template, does not preclude its use on cDNA from infection experiments.

3.5 Plate set-up

A standard total volume for each reaction is 20 μ l. To minimize costs, and when a good procedure has been established, this may be reduced further to half volumes or less. Each reaction contains.

- 1) 2X reaction buffer (10 μ l)
- 2) 20X Gene expression mix (1 μ l expression mix or 1 μ l each of different primers and 1 μ l of SYBR green)
- 3) Template (variable)

- i) cDNA (10ng in 7-9 μ l water)
 - ii) genomic DNA (in 7-9 μ l water)
 - iii) negative control (water only)
- 4) Nuclease free water to total volume 20 μ l

One way to set up a large plate with several assays is to make 2 sets of master mixes. First create a probe/primer or primer/SYBR green master mix (total volume 11 or 13 μ l respectively depending on whether SYBR green or primer/probes are being used). Add this to all of the wells/tubes and then add the Template, water and mix, where Template is at correct concentration (e.g. 10 ng in 9 or 7 μ l of water). In other commercial mixes, the Master Mix solution is at 2X with all the reagents including SYBR green, in which case 10 μ l of mix is added to the template which contains 1-10 ng of cDNA, the primers and water to a total volume of 20 μ l.

3.6 Running the reaction

Analysis will be easier if the plate is set up correctly. Although the details vary from machine to machine, the following commonalities apply. Each transcript assayed should be assigned an individual detector. Essentially even if FAM or SYBR green is being used for each transcript, you designate them separate 'detectors' as this allows separate analysis of each transcript. This is particularly useful for example if the 16S needs to be analyzed using different settings. If you are performing relative quantification the

endogenous controls should be specified and if the experiment has multiple samples then there should be an endogenous control for each sample (see above and Fig. 2).

Cycle conditions are typically XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX, with fluorescent measurements taken at the end of each cycle. If using SYBR green perform a melting curve analysis at the end of the cycling to show if more than one product has been amplified.

3.7 Analysing the data

The first step of any real-time PCR analysis is to adjust the baseline and threshold value (Figure 1). Generally the automatically generated baseline will be optimal but if not it should be adjusted as necessary (e.g. if plates are to be compared and the standards vary between plates that contain identical sample)s. The baseline should be set such that amplification occurs just after the maximum baseline value for all of the samples (see Figure 1). The threshold or C_T values should be adjusted such that it occurs in the early or middle exponential phase of the amplification curve. When these have been decided, it is necessary to reanalyze the data. Any C_T values which are obviously in error can be removed from the analysis (for example, if two C_T values are 15 and 1 is 45 – one could trust the two readings which are in agreement). After an appropriate baseline and C_T value have been chosen, the results can be analysed and exported either to a Microsoft Excel spread sheet or to custom real-time analysis software (often this is software unique to the light-cycler).

There are many ways in which to analyze both absolute and relative real-time PCR data and excellent help with dozens of useful references can be found on a website written and maintained by Michael Pfaffl (<http://www.gene-quantification.info/>). It is outside the scope of this chapter to exhaustively review analysis approaches, but for reviews on approaches to data analysis see (4, 12, 13).

Although the method of analysis should be decided upon before designing the experiment, if uncertain it is often best to add appropriate controls so as not to exclude future analysis possibilities. The setup shown in Table 1 contains enough references within the 96 well plate to be analyzed in several ways. Thus it could be normalized to the 16S using the $2^{\Delta\Delta CT}$ method (see below) or analysed using a modification of the method detailed below. Rather than exhaustively review analysis approaches, a specific example of analysis is given which I have found to be particularly useful. It is based on a modification of the $2^{\Delta\Delta CT}$ method and allows data to be presented in terms of number of phage transcripts.

In the standard $2^{\Delta\Delta CT}$ method, the amount of transcript in a sample is compared to an endogenous control and then to a calibrator (14). For this calculation to be valid, the amplification efficiency of the target (gene of interest) and the calibrator must be approximately the same. This can be determined by performing serial dilutions of template for both genes as described in section 3.3. The CT values at each dilution can then be subtracted from each other as a ΔC_T for each concentration and plotted on the y

axis against the log of the input amount of total genomic DNA. The slope of the graph should be <0.1 (14). Once even amplification efficiency has been established, at all time points, the C_T value of the endogenous control (in this case 16S at each time point) is subtracted from the time point to give ΔC_T and then $\Delta\Delta C_T$ is the result of normalizing this ΔC_T value to a calibrator (e.g. time point 0) which is obtained by subtracting the C_T value of the endogenous control at the calibrating time point. So for the set up shown in Table 1, the $\Delta\Delta C_T$ value at 3 hours for phage gene D1 relative to time point zero, would be the C_T value for 3 hours minus the 16S value for 3 hours, minus (the C_T value for phage D1 time point zero – the C_T value for 16S time point zero). In summary, when this value is plotted as $2^{-\Delta\Delta C_T}$ then gives the amount of target transcript (phage D1 at 3 hours) normalized to an endogenous control (16S at 3 hours) and relative to a calibrator (phage D1 at 0 hours – 16S at 3 hours). In this approach the standard deviations are calculated by taking the square root of the average calibrator value subtracted from the average sample value at any time point.

To measure the expression of phage and host genes during the infection cycle I modified this protocol in 2 ways: 1) for a calibrator I used purified phage genomic DNA (this allowed the data to be presented in terms of transcript number), and (2) to this end, in order to maintain absolute values of the transcripts being assayed the standard deviation of the 16S mean C_T value was subtracted from the transcript assay C_T value (see note 4).

An example template for Microsoft Excel of the modified analysis is shown in Table 2. The sample data shown are for the phage S-PM2 gene g47, in T4, gp47 is involved in

recombination. Column 2 shows the average C_T value g47 at each time point (see note 5). Column 3 shows the average 16 S C_T value subtracted from the actual 16S value for each time point (see note 4). The ΔC_T was therefore obtained by subtracting this ($C_{T_{16S}} - \text{average } C_{T_{16S}}$) from the gene 47 C_T value for each transcript (this is given in column 4). The calibrator (column 5) was the average C_T value of 4 phage genes (genomic template performed in triplicate) amplified from 10 ng of purified phage genomic DNA. Column 6 is the $\Delta\Delta C_T$ value obtained by subtracting column 5 from column 4. The next column (7) converts this value into the $2^{\Delta\Delta C_T}$. The inverse of this value is given in column 8 which is the amount of transcript at the various time points relative to 16S rRNA and 10 ng of phage DNA. This value is then divided by 10 to account for the fact that 10 ng of phage DNA was used as template. Finally the value in 9 is multiplied by 4,990,000 which is the number of copies of S-PM2 in 1 ng of DNA (see note 3). Therefore, by plotting this value against the time points we now have a meaningful and useful way of looking at the number of copies of individual phage transcripts (see note 6).

4 Notes

- 1) Unless absolutely necessary, duplexing reactions are generally not necessary.
- 2) Some master mixes contain UNG (Uracil-N-glycosylase which used in conjunction with dUTP prevents carry over contamination). UNG catalyzes the removal of uracil from uracil-containing DNA and thus ensures that any starting template is free from contamination of previously amplified products and generally is not necessary if good laboratory practice is maintained.

- 3) The number of phage genomes in the starting template of DNA used for real-time PCR can be calculated by dividing the ng of template used by the molecular mass of the phage and multiplying this value by Avogadro's number. To give an example for the scenario outlined in Table 3, if 1 ng of phage DNA are used as template and the phage has a genome size of 196,280, and the average molecular weight of a base pair of DNA is 615 Da, then the number of phage genomes = $[2 \times 10^{-9} / (196,280 \times 615)] (6.10^{23}) = 4.99 \times 10^6$. NB the reason this value is 2×10^{-9} is because DNA is double stranded and thus twice the number of genomes are counted as one genome equivalent is amplified from each strand of DNA.
- 4) This is simply to use the phage genomic control as a calibrator and convert the value into absolute number of phage transcripts.
- 5) The C_T value at time point 0 is 0 as no phages had been added at this point. This time point is not included in the table.
- 6) For this approach to work, all phage genes had to amplify equally efficiently at all C_T values. This proved to be the case and the standard deviation for all phage genes for a given template was negligible.

1. Heid CA, Stevens J, Livak KJ & PM., W. (1996) *Genome Research* **6**, 986-994.
2. Luke, K., Radek, A., Liu, X. P., Campbell, J., Uzan, M., Haselkorn, R. & Kogan, Y. (2002) *Virology* **299**, 182-191.
3. Poranen, M. M., Ravantti, J. J., Grahn, A. M., Gupta, R., Auvinen, P. & Bamford, D. H. (2006) *J. Virol.* **80**, 8081-8088.
4. Bustin, S. A. (2002) *Journal of Molecular Endocrinology* **29**, 23-39.
5. Dorak, M. T. (2006) *Real-time PCR* (Taylor and Francis.
6. Casas, V., Miyake, J., Balsley, H., Roark, J., Telles, S., Leeds, S., Zurita, I., Breitbart, M., Bartlett, D., Azam, F. & Rohwer, F. (2006) *Federation of European Microbiologica Societies: Microbiology Letters* **261**, 141-149.
7. Sandaa, R.-A. & Larsen, A. (2006) *Applied and Environmental Microbiology* **72**, 4610-4618.
8. Bookout, A. L. & Mangelsdorf, D. J. (2003) *Nuclear Receptor Signaling*, doi: 10.1621/nrs.01012.
9. Miller, E. S., Kutter, E., Mosig, G., Arisaka, F., Kunisawa, T. & Rüger, W. (2003) *Microbiology and Molecular Biology Reviews* **67**, 86-156.
10. Clokie, M. R. J., Shan, J., Bailey, S., Jia, Y., Krisch, H. M., West, S. & Mann, N. H. (2006) *Environmental Microbiology* **8**, 827-835.
11. Mann, N. H., Clokie, M. R. J., Millard, A., Cook, A., Wilson, W. H., Wheatley, P. J., Letarov, A. & Krisch, H. M. (2005) *Journal of Bacteriology* **187**, 3188-3200.
12. Rebrikov, D. V. & Tofimov, D. Y. (2006) *Applied Biochemistry and Microbiology* **42**, 455-463.
13. Pfaffl, M. W. (2001) *Nucleic Acids Research* **29**, 2000-2007.
14. (2001) *User bulletin #2. ABI Prism 7700 Sequence Detection System*. (Applied Biosystems) <http://docs.appliedbiosystems.com/pebi docs/04303859.pdf>.

Table 1. A typical RT-PCR plate set up. The numbers 0, 1, 3, 6 and 9 represent time after infection. D1h is an assay that detects both alleles of host *psbA*, D1I1 and D1I2 only detect one allele, D1p is an assay that detects bacteriophage encoded DNA, 16S is an endogenous control, RegA, gene 44, 47 and 18 are both fluorescent controls and calibrators. An endogenous control is present for each sample point, and a negative for each assay.

	1	2	3	4	5	6	7	8	9	10	11	12
A	0 A D1h	0 A D1h	0 A D1h	1 A D1h	1 A D1h	1 A D1h	3 A D1h	3 A D1h	3 A D1h	6 A D1h	6 A D1h	6 A D1h
B	9 A D1h	9 A D1h	9 A D1h	0 A D1p	0 A D1p	0 A D1p	1 A D1p	1 A D1p	1 A D1p	3 A D1p	3 A D1p	3 A D1p
C	6 A D1p	6 A D1p	6 A D1p	9 A D1p	9 A D1p	9 A D1p	0 A D1I1	0 A D1I1	0 A D1I1	1 A D1I1	1 A D1I1	1 A D1I1
D	3 A D1h	3 A D1h	3 A D1h	6 A D1h	6 A D1h	6 A D1h	9 A D1h	9 A D1h	9 A D1h	0 A D1I2	0 A D1I2	0 A D1I2
E	1 A D1I2	1 A D1I2	1 A D1I2	3 A D1I2	3 A D1I2	3 A D1I2	6 A D1I2	6 A D1I2	6 A D1I2	9 A D1I2	9 A D1I2	9 A D1I2
F	0 A 16S	0 A 16S	0 A 16S	1 A 16S	1 A 16S	1 A 16S	3 A 16S	3 A 16S	3 A 16S	6 A 16S	6 A 16S	6 A 16S
G	RegA 10ng DNA	RegA 10ng DNA	RegA 10ng DNA	44 10ng DNA	44 10ng DNA	44 10ng DNA	47 10ng DNA	47 10ng DNA	47 10ng DNA	18 10 ng DNA	18 10 ng DNA	18 10 ng DNA
H	9 A 16S	9 A 16S	9 A 16S	Neg RegA	Neg 44	Neg 47	Neg 18	Neg D1p	Neg D1h	Neg D1I1	Neg D1I2	Neg 16S

Table 2. An example of an Excel spread sheet designed to calculate the number of phage *psbA* (termed ‘D1p’ above) transcripts present at different infection times.

1	2	3	4	5	6	7	8	9	10
Time	C _T (D1p)	Control (16S correction)	Corrected ΔC_T (2)-(3)	Calibrator C _{T Ave}	$\Delta\Delta C_T$ (4)– (5)	2 ⁽⁶⁾	Relative transcript copy 1/(7)	8/10	Absolute transcript copy x phage
1	20.058	-0.160	20.218	14.362	5.856	57.925	0.017263	0.0017263	8614.587
3	17.993	-0.227	18.220	14.362	3.858	14.501	0.068959	0.0068959	34410.61
6	19.798	0.429	19.369	14.362	5.007	32.158	0.031096	0.0031096	15517.04
9	20.089	-0.278	20.367	14.362	6.005	64.227	0.015570	0.0015570	7769.284

Figure 1. Typical amplification curves from 4 samples of cDNA. Note amplification begins shortly above the baseline, and the C_T or threshold value is set during the exponential phase of the reaction. The point at which the C_T value is calculated is shown for the amplification curve at the far right of the curves. The X axis shows the cycle number, the Y axis Relative Fluorescence and ΔR_n vs Cycle = Δ

Figure 1