

PCR and partial sequencing of bacteriophage genomes

Martha Clokie

Dr Martha Clokie
Lecturer in Microbiology
Department of Infection, Immunity and Inflammation
Medical Sciences Building
University of Leicester
PO Box 138
Leicester LE1 9HN, U. K.

Tel: 0116-2522959 (Office)
Fax: 0116-2525030
email: mrjc1@le.ac.uk

Running title: PCR and sequencing

Key words: PCR, amplification, template, DNA, sequencing, primer

Summary/Abstract

PCR is a quick and effective way of identifying the presence and ‘affiliation’ of bacteriophages, or phage encoded genes from environmental samples, bacterial cells or purified viruses. The limitations are that you have to know what you are looking for in order to find it. Although the bacteriophage world doesn’t have the advantage of a conserved gene, present in all members, there are many phage genes that do show nucleotide conservation even between phage which infect fairly divergent taxa. As more sequence data becomes available both through metagenomic approaches and the sequencing of complete bacteriophage genomes, PCR primers can be increasingly refined and thus it should be an increasingly useful tool for bacteriophage biology.

1. Introduction

This chapter is not designed to be a panacea to all PCR and sequencing related problems in bacteriophage biology however it does systematically cover the main phage specific issues regarding the techniques. For the PCR neophyte or even beginner, the following website has very useful information <http://www.horizonpress.com/pcr/>.

To characterise a newly isolated bacterial strain, one of the first tools used is the polymorphism of a universal marker, typically the sequence of the 16S ribosomal DNA. This gives information of the phylogenetic position of the bacterial strain. All bacteria have ribosomes and thus ribosomal genes which fortunately make good markers due to having highly conserved regions, thus ‘universal’ primers can be used to amplify them and highly variable regions which can be used to distinguish between bacterial strains.

Unfortunately bacteriophages have no such gene that is common to all representatives. Even when a gene may be common to all three phage families e.g. the major structural genes, the sequences are so varied that an alignment and consequently a phylogenetic analysis is not possible (Chapter 31). One method to estimate viral diversity is to use metagenomic sequencing (Chapter 23) followed by custom designed arrays to probe known viral sequences in a range of environments. However this scale of project is still expensive in terms of labour and cost. Often it is useful to just get an idea of what family a virus belongs to, or whether it possesses a particular gene of interest. Alternatively one may be investigating the presence of phages or phage-encoded genes as prophages or in bacterial communities. For example phage encoded toxins contributing to pathogenesis in cholera, diphtheria, enterohemorrhagic diarrhea and *Staphylococcus aureus* were successfully screened for using PCR based approaches (1). For many small scale focused projects PCR is still the preferred technique.

The first step in bacteriophage PCR is obtaining suitable high-quality template from which to amplify genes. Fortunately in most cases PCR requires very little DNA (~1 – 10 ng of phage DNA per reaction), however there are cases where more is required (such as direct sequencing). DNA may be extracted from environmental samples (Chapter 2), phage lysate prepared from either a liquid culture or scraped plates, (Chapter 23), infected cells or uninfected cells if prophages are being looked for. cDNA can also be a good template for PCR, in which case good quality RNA has to be extracted either from infected cells or isolated from a bacterial community

(Chapter 34). The template can also be diluted high titre phage lysate or removed from community DNA extracted from agarose gels.

One method of increasing the amount template for downstream applications such as sequencing is via the use of phage derived enzymes e.g. Genomiphi from GE Healthcare. This enzyme (the DNA polymerase from *Bacillus* phage $\phi 29$) is particularly useful for samples where it is difficult to obtain large quantities of DNA since it converts nanograms of DNA to micrograms of DNA overnight. The product can then be suitably used for direct sequencing, PCR or library preparation (Chapter 26). Direct sequencing is particularly useful for example where the presence of a gene in a novel genome has been detected using PCR, but it isn't know what genes lie to either side. It is therefore possible to design primers to walk out from the known gene and to walk out by using the total amplified phage genome as a template (2).

The biggest challenge in bacteriophage PCR and sequencing is the choice of gene to be amplified and the consequent design of primers. The gene of choice is obviously dependent upon the question being addressed. For questions of diversity the best understood bacteriophage 'genus' is the 'T4 – like' viruses in the family *Myoviridae* where primers have been designed for the major capsid proteins gp23 and gp20 and for the gene which encodes DNA polymerase (gp43). These have been predominantly used to study diversity in cyanophages (3-5) but they have also been effective in oceanic viruses in general (6). Caution must always be taken however in the interpretation of authors' claims to virus primers being 'universal' as they may not be ubiquitously so. Even genes which are common to all 'T4 – like' myoviruses are widely divergent (7).

No such environmental screens have been carried out for oceanic *Siphoviridae* and *Podoviridae*, likely reflects the paucity of knowledge of these phage sequences. As more sequence data becomes available both through sequencing metagenomes and individual phage genomes, probing the 'phageome' with PCR based approaches will become easier. Similarly viruses which infect archaea have been less studied in many ways than phages of eubacteria and due to the high diversity of genes that have been identified, no such environmental screens of diversity based on PCR have been attempted. When sufficient study has been performed on a group of phages in a particular environment, sequence data can be generated which then lends itself to further study by PCR. An example of this is in the dairy industry where bacteriophages are a major economic problem. In *Lactococcus* most phages belong to one of three major groups of *Siphoviridae* (8). This allows multiplex PCR primers to be designed for a relatively rapid screen of dairy plants. However recent more detailed analysis has identified many novel groups even within this narrow environment that cannot be detected with PCR (9).

When working with individual bacteriophages then electron microscopy can be used to establish the family to which they belong (Chapter 12). Then at least it is possible to establish for example whether the phage was a myovirus and whether to expect a gp20 product from a PCR reaction. Although the morphology of viruses does not always reflect their genetics, if the 'genus' within the phage family can be established (Chapter 12) there is a higher chance of perhaps using this information to design appropriate primers.

If the aim of the project is to identify temperate bacteriophages then primers specific to integrases may be a suitable target although it is worth remembering how varied such sequences can be (10). The identity of the bacteriophage family is not so useful when screening for bacteriophage encoded genes acquired from their bacterial hosts. For example cyanophage acquired photosynthesis genes may be amplified equally successfully with the same primers from either powdered viruses or myoviruses (11). If no sequence data is available for the bacteriophage/s being studied then it may be possible to design primers from closely related bacteriophages. However as sequences can be so varied, even with significant degeneracy often this may be of little use. In an ideal world one could completely sequence the entire genome of the bacteriophage (Chapter 26) however in reality this is currently outwith the scope of most projects and budgets. The presence or absence of particular genes may be determined using Southern blotting. Cold and radiolabeled RFLP's (Chapter 26) may be done in parallel and the appropriate fragment of gene from the cold version can be cloned and sequenced. If more than one bacteriophage has been isolated for a new organism, it may be useful to determine the genes they have in common as if the phages are in the same family, they are likely to share structural genes. It may be the case that even hybridisation approaches fail and one solution is to take proteomic approach to identify the major structural proteins (Chapter 18 and 39). This will allow degenerate primers to be designed and sequence data generated. Although this approach may seem cumbersome, it has been successful when trying to obtain sequence data for poorly characterised bacterial taxa (12).

2. Materials

All solutions, plastic and glass wear and equipment should be clean and sterile. PCR reagents and DNA should be kept at -20 °C. All other solutions should be kept at room temperature unless stated otherwise.

2.1 Target DNA (template) for PCR

This is obtained as described below.

2.2 Target DNA (template) for sequencing

1. Nanogram quantities of DNA template.
2. Commercially available kit for amplifying circular DNA e.g. Genomiphi (GE Healthcare; <http://www.gehealthcare.com/> formerly Amersham Bioscience)

2.3 Primers

These can be purchased from any oligo synthesising unit – e.g. Sigma Genosys (The Woodlands, Texas; http://www.sigmaaldrich.com/Brands/Sigma_Genosys.html), Invitrogen Corp. (Carlsbad, CA; <http://www.invitrogen.com/>) and many small companies.

2.4 General Reagents for PCR

1. Ultra-pure H₂O to 50 µl
2. PCR polymerase enzyme with corresponding PCR reaction buffer (commercially available)
3. 5 µl 2mM dNTP solution (commercially available from PCR enzyme producer).
4. Primers: 10 µM working solution in ultra-pure H₂O (*see Note 1*)
5. 1 ng – 1 µg/µL Template DNA

2.6 General reagents for Gel electrophoresis visualisation

1. Agarose
2. Ethidium bromide (50 µg/ml)

3. Loading buffer (6X concentration, to make 100 ml): add 93.6 ml of glycerol to 153.4 ml of water, 3 ml of 0.3 M EDTA, 0.3 g of bromophenol blue and 0.3 g of xylene cyanol
4. TAE running buffer (50X stock solution): 24.2% (w/v) Tris-HCl pH 7.5, 5.71% (w/v) acetic acid, 3.72% (w/v) EDTA.2H₂O (adjust to pH 8 with HCl).
5. DNA size concentration ladder (e.g. New England Biolabs; Ipswich, MA; <http://www.neb.com/> or Fermentas Inc; Hanover, MD; <http://www.fermentas.com/>)

2.7 Equipment

1. Thermocycler or PCR machine. Any make will do. If using an established protocol make sure that ramping times are consistent.
2. Incubator or water bath at 37 °C to perform the overnight incubation when amplifying whole phage genomes.
3. Spectrophotometer to accurately quantify DNA or RNA template. Use quartz cuvettes if quantifying nucleic acid in a spectrophotometer that requires them. A NanoDrop spectrophotometer can be especially useful if one has only limited amounts of template (NanoDrop Technologies; Wilmington, DE; <http://www.nanodrop.com/>)
4. Gel documentation system or UV transilluminator and camera

3. Methods

3.1 Template preparation

A 1 µl sample of a 1:5 dilution of a $\sim 10^{11}$ phage stock diluted in dH₂O may produce a suitable template without having to prepare pure DNA through extraction. An alternative method is to pick a phage plaque and re-suspend it in 50 µl of water. Leave at room temperature for 30-60 min and then boil for a few minutes in a water bath (13). Up to 25 µl of this cleaned phage lysate may be necessary for PCR. (*see*

Note 2). DNA may also be extracted from PFGE experiments by performing two rounds of freezing and thawing of the plugs or bands of interest followed by centrifugation at 1500 x g to sediment the agarose. The supernatant can be directly used as a template for PCR. This type of approach is often useful when probing community viral DNA, following PFGE analysis to look for the presence of particular genes (14).

If these approaches do not work then a DNA extraction has to be carried out. See Chapters 2, 23 and 34. Generally 1-100 ng of high quality DNA template is necessary per reaction. As for any DNA, extracted template should be stored in either ultra pure water or in Tris buffer (pH 7). For reference to appropriate chapters to produce suitable template please see above.

3.2 Target DNA (template) for sequencing

Sequencing of phage genes may be performed from PCR products in the usual way. Alternatively they may be sequenced directly from phage DNA prepared from Genomiphi (GE Healthcare) or a similar enzyme. This is useful if either it is not possible to get a PCR product to work, or it is necessary to walk out from a known sequence to genes that are not known. Refer to manual for instructions. In brief the template is denatured and then left at 37 °C overnight in the presence of the DNA polymerase from ϕ 29 in the case of Genomiphi and random hexamers. The reaction is denatured by heating at 60 °C for 5 min. The resulting DNA is quantified by absorbance at 260 °C, 1 µg of DNA per sequencing run is required.

3.3 Primer design

Primers are designed in the usual way and it is a good idea to use software to design them to avoid problems with dimmers, runs and hairpins and appropriate GC contents and melting temperatures. For information on software, see <http://molbiol-tools.ca/PCR.htm>. The better the primers, the fewer downstream problems there are with PCR. Commonly used primer sequences for diversity screening of marine bacteriophages are given in Table 1. These give the exact sequences for cyanophage specific primers from the *Myoviridae* (5, 15-17) and for more general members of the T4 type phages (6). The primers also illustrate the amount of degeneracy required even in these conserved genes in order to detect as many different isolates as possible.

Table 1 to show commonly used primers in exploring myoviruses diversity

Phage gene	Primer direction and name	Sequence 5' – 3'	Reference:
Gp 20	F CPS1 R CPS2	GTAG(T/A)ATTTTCTACATTGA(C/T)GTTGG GGTA(G/A)CCAGAAATC(C/T)TC(C/A)AGCAT	(3)
Gp20	F CPS3 R CPS4 R CPS8	TGGTA(T/C)GT(T/C)GATGG(A/C)AGA CAT(A/T)TC(A/T)TCCCA(A/T/C)TCTTC AAATA(C/T)TT(G/A/T)CCAACA(A/T)ATGGA	(17)
Gp20	R G20-2	(G/C)(A/T)(A/G)AAATA(C/T)TTICC(A/G)AC(A/G) (A/T)A(G/T)GGATC	Short and Suttle
Gp23	Mzia 1 Mzia 2	GATATTTGIGGIGTTCAGCCATGA CGCGGTTGATTTCAGCATGATTTC	Filee et al, 2005 (6)

Sequences from which to design primers are obtained from GenBank. For general sequence recovery information on how to do this see <http://molbiol-tools.ca/GenBank.htm>.

3.3 Amplification reaction

There is nothing different about amplification conditions for amplifying genes from bacteriophage DNA as compared with other template. In brief therefore, denature the DNA template for 5 min at 95 °C. Perform 20-30 cycles of denaturation at 95 °C (30 sec), annealing at the predicted melting temperature of the primer (30 sec) (*see Note 3*), and extension at 72 °C for (the time here depends on the product length to be amplified but a general rule of thumb is 30 seconds per 500 bp). Perform a final elongation at 72 °C for 5 minutes and store at 4 °C (*see Note 4*).

3.4 PCR Enhancers

A number of additives can be included to encourage PCR to work more efficiently. These can be particularly helpful when DNA has been obtained from environmental sources that contain inhibitory substances such as from blood, faeces, soil or aqueous environments. Useful additives include DMSO (dimethyl sulfoxide), betaine (N,N,N-trimethylglycine = [carboxymethyl]trimethylammonium), Formamide, Non-ionic detergents such as Triton X-100, Tween 20 or Nonidet P-40 (NP-40), TMAC (tetramethylammonium chloride), 7-deaza-2'-deoxyguanosine (dC⁷GTP), BSA (bovine serum albumin), and the T4 gene 32 protein. These inhibitors generally influence the structure of DNA, reducing secondary structure and allowing the template to be more amenable to the PCR. They are all useful in specific circumstances, and if used incorrectly can do more harm than good. Further information can be found <http://www.staff.uni-mainz.de/lieb/additiva.html>.

3.5 Detection and analysis of reaction product

Again there is nothing specifically phage related to be done here. Make a 0.8 – 4% agarose gel in 1X TAE buffer. The percentage of agarose depends on the size of the product. A 500 bp – 1 kb fragment will be perfectly visualised in a 1% gel. Weigh out the appropriate amount of agarose and add to the buffer. Melt in the microwave for around 2 min depending on the amount needed. Cool the gel either by swirling under a cold water tap or on the bench. Add 5 µL/50 mL of ethidium bromide and pour into casting tray.

Load around 5 µl of product from the PCR reaction. Run the amplified fragment alongside DNA ladder. Depending on the size of the gel run between 60-240 volts through it until the dye front has migrated at least 5 cm.

4. Notes

1. The primers usually arrive lyophilised and the amount in ng (or nmoles) synthesised is given in the paper work. Add 1 ml of ultra-pure water to the lyophilised primer and to calculate a working concentration of 10 µM use the equation $c_1v_1 = c_2v_2$, where c is concentration and v is volume. For example if you have added 1 ml of water to 50 ng of primer, to make the 10 mM working stock in 100 µL you will need to add 20 µL of concentrated primer and 80 µL of ultra pure water.

2. Adding too much DNA can result in non-specific amplification. When amplifying from bacterial DNA when screening for phage products or phage encoded products more DNA may be required (up to 1000 ng). Care must be taken to limit the amount of Taq polymerase inhibitors as much as possible (such as detergent, EDTA and

traces of phenol/chloroform). Therefore diluting DNA template can result in a successful reaction as it may dilute out an inhibitors present.

3. This is can be predicted by in the software package that you design your primers in or a calculator is available from <http://www.promega.com/biomath/default.htm>, the information will also be included on the details when the primers are delivered.

4. If the PCR is set up overnight, do not leave on 4 °C until morning as this puts excessive strain on the PCR machine and will significantly shorten its life.

1. Casas, V., Miyake, J., Balsley, H., Roark, J., Telles, S., Leeds, S., Zurita, I., Breitbart, M., Bartlett, D., Azam, F. & Rohwer, F. (2006) *FEMS Microbiology Letters* **261**, 141-149.
2. Millard, A., Clokie, M. R. J., Shub, D. A. & Mann, N. H. (2004) *Proceedings of the National Academy of Sciences of the United States of America* **101**, 11007-11012.
3. Fuller, N. J., Wilson, W. H., Joint, I. R. & Mann, N. H. (1998) *Applied and Environmental Microbiology* **64**, 2051-2060.
4. Marston, M. F. & Sallee, J. L. (2003) *Appl Environ Microbiol* **69**, 4639-47.
5. Short, C. M. & Suttle, C. A. (2005) *Applied and Environmental Microbiology* **71**, 480-486.
6. Filee, J., Tetart, F., Suttle, C. A. & Krisch, H. M. (2005) *PNAS* **102**, 12471-12476.
7. Petrov, V. M., Nolan, J. M., Bertrand, C., Levy, D., Desplats, C., Krisch, H. M. & Karam, J. D. (2006) *Journal of Molecular Biology* **361**, 46.
8. Miklic, A. & Rogelj, I. (2003) *International Journal Of Food Science and Technology* **38**, 305-311.
9. Deveau, H., Labrie, S. J., Chopin, M.-C. & Moineau, S. (2006) *Appl. Environ. Microbiol.* **72**, 4338-4346.
10. Balding, C., Bromley, S. A., Pickup, R. W. & Saunders, J. R. (2005) *Environmental Microbiology* **7**, 1558-1567.
11. Sullivan, M. B., Lindell, D., Lee JA, Thompson, L. R., Bielawski, J. P. & Chisholm, S. W. (2006) *PLOS* **4**, 1344-1357.
12. Baker, A. C., Goddard, V. J., Davy, J., Schroeder, D. C., Adams, D. G. & Wilson, W. H. (2006) *Appl. Environ. Microbiol.* **72**, 5713-5719.
13. Kutter, E. & Sulakvelidze, A. (2005) *Bacteriophages: Biology and Applications* (CRC Press, Boca Raton).
14. Sandaa, R.-A. & Larsen, A. (2006) *Appl. Environ. Microbiol.* **72**, 4610-4618.
15. Fuller, N. J., Marie, D., Partensky, F., Vaulot, D., Post, A. F. & Scanlan, D. J. (2003) *Applied and Environmental Microbiology* **69**, 2430-2443.
16. Wilson, W. H., Fuller, N. J., Joint, I. R. & Mann, N. H. (1999) *Bulletin de l'Institut océanographique, Monaco* **19**, 209-216.
17. Zhong, Y., Chen, F., Wilhelm, S. W., Poorvin, L. & Hodson, R. E. (2002) *Applied and Environmental Microbiology* **68**, 1576-1584.

