Use of non-linear mixed-effects modelling and regression analysis to predict the number of somatic coliphages by plaque enumeration after 3 hours of incubation

Javier Mendez, Antonio Monleon-Getino, Juan Jofre and Francisco Lucena

ABSTRACT

The present study aimed to establish the kinetics of the appearance of coliphage plaques using the double agar layer titration technique to evaluate the feasibility of using traditional coliphage plaque forming unit (PFU) enumeration as a rapid quantification method. Repeated measurements of the appearance of plaques of coliphages titrated according to ISO 10705-2 at different times were analysed using non-linear mixed-effects regression to determine the most suitable model of their appearance kinetics. Although this model is adequate, to simplify its applicability two linear models were developed to predict the numbers of coliphages reliably, using the PFU counts as determined by the ISO after only 3 hours of incubation. One linear model, when the number of plaques detected was between 4 and 26 PFU after 3 hours, had a linear fit of: $(1.48 \times \text{Counts}_{3\,h} + 1.97)$; and the other, values >26 PFU, had a fit of $(1.18 \times \text{Counts}_{3\,h} + 2.95)$. If the number of plaques detected was <4 PFU after 3 hours, we recommend incubation for (18 ± 3) hours. The study indicates that the traditional coliphage plating technique has a reasonable potential to provide results in a single working day without the need to invest in additional laboratory equipment.

Key words | coliphages, prediction, probability, rapid detection, regression modelling

Javier Mendez (corresponding author)
Juan Jofre
Francisco Lucena
Department of Genetics, Microbiology and Statistics, Faculty of Biology, Prevosti Building,
University of Barcelona,
Av. Diagonal 643,
08028 Barcelona,
Spain
E-mail: *jmendez@ub.edu*

Antonio Monleon-Getino
Research Group in Biostatistics and Bioinformatics (GRBIO), Department of Genetics, Microbiology and Statistics, Faculty of Biology, Aulari Building,
University of Barcelona,
Av. Diagonal 643,
08028 Barcelona,
Spain

ABBREVIATIONS

NLME non-linear mixed-effects model
AIC Akaike's information criterion
BIC Schwarz's Bayesian information criterion

INTRODUCTION

The transmission of water-borne diseases is still a matter of major concern, despite worldwide efforts and the modern technology used for sanitation and the production of safe drinking water (WHO 2008). To protect humans against possible water-borne diseases, ideally the occurrence and levels of all pathogens in waters for different uses, particularly for drinking, should be monitored. This ideal is, nevertheless, not realistic for a number of practical and economic reasons. The main procedural problems are due to low numbers, unequal distributions through the year and, frequently, extremely complex methods; the economic drawbacks are the need for expensive reagents, sophisticated equipment and expertise. To overcome these problems, the use of indicator organisms that are readily detected (indexes) has been widely adopted to assess either the microbiological quality of water (drinking water, source water, bathing water, reclaimed water, etc.) or the performance of water treatment processes (Ashbolt *et al.* 2001). At present, the most commonly used indicator

doi: 10.2166/wh.2017.067

microorganisms are bacteria (coliform bacteria, faecal coliform bacteria, *Escherichia coli*, enterococci, and spores of sulphite reducing clostridia). In addition, bacteriophages are increasingly being included in some control programmes. Indeed, several groups of bacteriophages, including somatic coliphages, have been suggested as indicator microorganisms for use in the assessment of faecal contamination of water (IAWPRC 1991; Armon & Kott 1996; Grabow 2001; Lucena & Jofre 2010) and food (Hsu *et al.* 2002; Pillai 2006).

Somatic coliphages are those phages that replicate in *E. coli* after infecting it through the cell wall (Kott *et al.* 1974; IAWPRC 1991) and they are not only valuable indexes of faecal pollution, but also indicators of the behaviour and survival of human viruses with which they share many basic features. They have gradually been added to water quality guidelines, such as the Ground Water Rule (USEPA 2006), the indications for the quality of drinking water in the Canadian state of Québec (Anonymous 2001), those for reclaimed water in Queensland, Australia (Anonymous 2005) and reclaimed water in some American states (USEPA 2012). Standardised methods for detecting and quantifying this group of phages have been established and are at present available (ISO 1995; USEPA 2001a, 2001b). The ISO and USEPA standard methods are very similar and provide identical counts of somatic coliphages (Guzmán *et al.* 2008).

However, whether testing for pathogens or for traditional indicator organisms, there is a common need for rapid analysis either to monitor the performance of treatment processes in order to adjust them when necessary, or to determine the microbiological quality of water available for consumers, either for drinking, bathing, irrigation or recharging aquifers. Too often, the results of testing are not available until long after the water has been used. Therefore, there is a need for methods with reduced response times that provide data in the shortest possible time. At present, the fastest standardised methods provide results in 18–24 hours, particularly when low numbers of bacterial indicators are present in the sample, as is the case of samples that need microbiological analysis (Venter *et al.* 2000; APHA 2001). Some indirect methods with reduced response times have been reported for the detection of *E. coli*. Moreover, the detection of specific enzyme activities such as β-galactosidase for total coliform bacteria or β-glucuronidase for *E. coli* can be shortened by using spectrofluorometry, indirect impedance or liquid scintillation spectroscopy. However, all of these methods have poor sensitivity for the ranges of values required in water analysis and need sophisticated apparatus not commonly available in standard laboratories, which are serious drawbacks for these methods (Reasoner & Geldreich 1989; Colquhoun *et al.* 1995; Park *et al.* 1995).

Meanwhile, molecular techniques such as real-time q-PCR allow the time needed to obtain results to be shorted and, at the same time, low amounts of contaminants to be detected. However, positive detection by these nucleic acid amplification methods are no assurance of infectivity (Sobsey *et al.* 1998; Gassilloud *et al.* 2003; Jofre & Blanch 2010), and infectivity is essential for determining the risk associated with the presence of infectious microbes and for monitoring treatment processes totally or partially based on disinfection.

Regarding somatic coliphages, the ISO standard method (ISO 1995) for detecting plaque forming units (PFUs) establishes 18 hours as the optimal incubation time, but it considers the possibility of counting PFUs after 6 hours. Here, we present data to compare the counts after different times of incubation and report a mathematical approach based on non-linear mixed-effects (NLME) modelling and regression analysis that provides a reliable prediction of the results in 3 hours, providing the same information as the standard method regarding both numbers and viability.

MATERIALS AND METHODS

Samples, sampling and sampling sites

River water

Ten river samples were obtained from the rivers Llobregat and Ter (NE Spain), collected in sterile 1,000 mL polyethylene containers, transported to the laboratory within 6 hours of collection and stored at (5 ± 3) °C (maximum storage time: overnight) until they were used. The samples were titrated undiluted and diluted 1:10. The 1:10 dilutions were prepared as indicated in the ISO standard using peptone

saline solution. From each sample, five replicates of 1 mL aliquots were titrated.

Raw municipal wastewater

Eleven raw sewage samples were obtained from two municipal wastewater treatment plants in Barcelona. Collection, transport and sample storage were the same as for the river samples. The samples were diluted 100-, 1,000- and 10,000-fold, and from each diluted sample three to five replicates of 1 mL aliquots were titrated.

Bacteriophage titration

PFUs of bacteriophages were counted using the double agar layer technique, according to ISO 10705-2. Plaques were recorded after 2, 3, 4, 5, 6 and 18 hours of incubation, without any sort of contrast or magnification at $(36 \pm 2)\ ^{\circ}C$ using *E. coli* WG5 as the somatic coliphage host. Counts at 2, 3, 4, 5 and 6 hours were carried out in a thermostatic room to avoid any effect of temperature change during the counting process.

Somatic coliphages are a heterogeneous group producing heterogeneous plaques that can differ in size and morphology. As plaque size tends to increase over the incubation time, plaques that are close together can overlap. Once the appearance of a plaque was recorded at the different times, the agar plate was counted again at 18 hours, according to ISO 10705-2. Plaque overlap was measured as the variation between the cumulative sums of the presence of new plaques recorded after the different incubation times contrasted with the single final count at 18 hours.

Quality assurance

To ensure the accuracy of bacteriophage titrations, phage stocks were prepared according to the ISO standard for the validation of methods for the concentration of bacteriophages from water (ISO 10705-3). Briefly, a primary or secondary sewage effluent was filtered through a 12 μm pore-size membrane filter. Target bacteriophages were enumerated by initial titre. The phage suspensions were stored at $(5 \pm 3)\ ^{\circ}C$ until the enumeration results were available. If necessary, the suspensions were diluted to obtain concentrations of 60 to 200 PFU/mL of each of the phage groups. After that, glycerol was added to a final concentration of 10% (v/v) and the phage suspensions were distributed into 10 mL volumes, in polypropylene tubes and frozen at $(-70 \pm 10)\ ^{\circ}C$. Before use, Cochran homogeneity tests were carried out on the vials to ensure the coliphage dilutions were homogeneous within and between the tubes (Mendez *et al.* 2002).

Controls of the host strain and the media were performed in duplicate for each bacteriophage enumeration series.

Coliphage plaque size determination

Briefly, a 20-megapixel digital camera was used to take plate images after 3 and 18 hours of incubation. The geometry of images were corrected to their real shape. The areas of the plaques were measured with the ImageJ software (Rasband & ImageJ, U. S. 1997–2016), to convert the pixel count to surface area; a computer printout with a known surface area was generated and it was used as the size standard. In this study, we found that 1 pixel $= 0.008\ mm^2$. Once the areas were determined the mean diameter of each plaque was calculated.

Data computation and statistics

Regression fits

Statistical analysis and modelling were carried out with the R statistical software package version 3.0.2 (R Development Core Team 2014) and locfit (Loader 2013).

Several functions including linear polynomic functions, and growth, sigmoidal, power-law and exponential models were tested using non-linear regression. This analysis was performed using the R-package nls2 (Grothendieck 2013).

Mixed-effects models

Traditionally, the variability associated with the detection of bacteriophage in water samples has been described using simple statistical terms such as the mean and standard deviation. In contrast, mixed-effects modelling allows us to include other sources of variability that exist within the

water samples, as well as the magnitude of the variability between repeated counts and the magnitude of the residual deviations between the predicted and observed values.

The non-linear plaque count curves were analysed by NLME modelling with the R-packages nlme (Pinheiro *et al.* 2015) and lme4 (Bates *et al.* 2015).

RESULTS

Coliphage counts

Coliphage counts were carried out without using any sort of contrast or magnification; 45 and 50 counts were performed, respectively, on the undiluted and 10-fold diluted river samples; 43 samples of sewage diluted at 100-, 1,000-, 10,000-fold were titrated, but in the case of the most diluted sewage samples, only 21 samples could be taken into account in this study due to the limited coliphage presence. The results are summarised in Table 1.

Determination of the plaque sizes at 3 and 18 hours

Diameters were determined measuring 176 coliphages plaques on five counts of undiluted river samples after their incubation at 3 and 18 hours A statistical difference ($p < 0.01$) between the diameters measured at 3 and 18 hours was observed when the Wilcoxon matched-pairs signed-rank test was applied. The mean diameters (mm) of the plaques measured at 3 hours was 1.81 mm and at 18 hours was 3.56 mm; the percentiles values at 0%, 25%, 50%, 75% and 100% were 0.4, 1.2, 1.6, 2.2 and 4.5 mm for 3 hours and 0.6, 1.4, 2.4, 5.2 and 14.4 mm for 18 hours.

Approximately 24.4% of the plaques did not experience an increase in diameter size, 55.7% increased less than 50%, 13.1% increased between 50% and 100% in size, 15.3% increased between 100% and 200%, 6.8% increased between 200% and 300% and 9.7% increased more than 300% in diameter size.

Overlapping plaques

The phages detected after 18 hours by counting at several incubation times was statistically higher than those detected at 18 hours by the single count procedure, Wilcoxon signed-rank test ($p < 0.05$). For that reason, the total 18-hour counts recorded at several incubation times were used in the statistical analysis.

It should be noted that, although the difference between the counting procedures was statistically significant, the

Table 1 | Overall coliphage titration results

Sample	Time (h)	n	Mean	SD	Min.	Max.	Sample	Time (h)	n	Mean	SD	Min.	Max.
Sewage (1:100)	2	43	181.23	38.68	96	268	River	2	45	53.47	24.69	10	124
	3	43	244.07	47.11	24	336		3	45	83.58	32.89	27	158
	4	43	266.91	48.80	149	352		4	45	97.80	35.38	37	176
	5	43	280.65	50.33	158	362		5	45	105.84	37.01	42	192
	6	43	292.67	52.94	162	384		6	45	112.18	37.66	44	195
	18	43	311.86	50.67	180	396		18	45	127.44	39.19	52	214
Sewage (1:1,000)	2	43	17.49	5.32	6	32	River (1:10)	2	50	4.14	2.89	0	11
	3	43	26.19	7.56	8	42		3	50	6.94	4.10	0	21
	4	43	30.12	9.09	11	48		4	50	8.36	4.53	0	23
	5	43	32.04	9.42	12	49		5	50	9.14	4.76	0	25
	6	43	34.35	10.14	13	50		6	50	9.76	4.77	2	26
	18	43	40.16	11.29	18	59		18	50	12.40	5.42	4	29
Sewage (1:10,000)	2	21	2.00	1.84	0	8							
	3	21	2.86	1.90	0	8							
	4	21	3.62	2.56	0	11							
	5	21	3.86	2.37	1	11							
	6	21	4.04	2.50	1	11							
	18	21	5.10	2.90	1	13							

differences were not very important: for plates containing 30 to 300 plaques, the average percentage difference was 2.8%, reducing to 2.9% when locally weighted smoothing was applied.

Figure 1 shows the results of the comparison between the count procedures. Locally weighted scatterplot smoothing (LOWESS) is a non-parametric strategy to fit a smooth best-fit curve without assuming the data must fit some distribution shape; a local regression model is fitted to each point and the points close to it.

Determination of minimum incubation time

As the aim of the study was to establish the feasibility of using early counting to predict the titration results that would be obtained in the standard method, our first approach was based on correlation analysis (Table 2) between the bacteriophage counts obtained at different times against the counts obtained at 18 hours according to the ISO procedure.

Although coliphages were detected after as little as 2 hours of incubation, bacteriophage counts at 3 hours were selected for modelling as they displayed more than 70% correlation in all the samples tested.

The correlations were notably weaker among samples containing low numbers of bacteriophages. For all cases, the correlations were significant at $p < 0.01$.

The worst condition corresponds to 10,000-fold diluted sewage samples, which presented an adjusted R^2 of 0.50 and a mean value of 5 PFU/mL, with coliphage counts ranging between 1 and 13 PFU/mL; whereas the best fit was achieved with undiluted river sample titrations: adjusted R^2 of 0.93, mean value of 117.90 PFU/mL, coliphage counts ranging between 33 and 214 PFU/mL.

NLME with random intercept modelling

Plaque appearance over the incubation period was curvilinear, as shown in Figure 2. Hence, several functions such as growth, sigmoidal, power-law, hyperbolic, rational and exponential functions were tested in order to find the best fit. Once they were determined, NLME analysis was carried out with each selected function.

Mixed effects can be applied for repeat measurements; the main difference in comparison with classical regression is that the observations may be grouped in independent levels or clusters, where observations in a level are dependent because they belong to the same subpopulation. Consequently, there are two sources of variation: between and within levels. Therefore, mixed-effects models provide the flexibility to model not only the mean of a response variable, but its covariance structure as well. As mixed-effects modelling introduces fixed and random effects over the

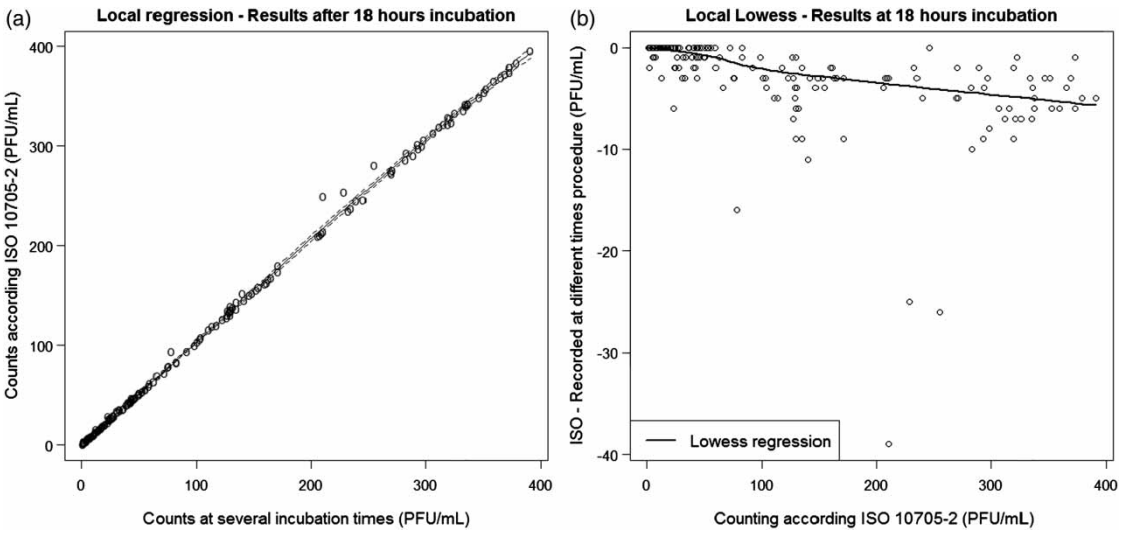

Figure 1 | Local regression between count procedures (a), the dashed line denotes no differences between counts and local LOWESS (b) smoothing (thick black line).

Table 2 | Pearson correlations for sewage and river samples

Pearson correlation ($count_{ih}$ v $count_{18\,h}$)

Hour	Sewage (1:100) r	Sewage (1:1,000) r	Sewage (1:10,000) r	River (undiluted) r	River (1:10) r
2	0.690	0.475	0.637	0.918	0.736
3	0.928	0.856	0.728	0.961	0.857
4	0.968	0.915	0.842	0.983	0.887
5	0.980	0.947	0.839	0.988	0.891
6	0.988	0.977	0.849	0.991	0.925
18	1.000	1.000	1.000	1.000	1.000

fitting variables, mixed models present multidimensionality and, for that reason, randomised variables have a matrix representation.

The goodness of fit (GOF) of a statistical model describes how well it fits a set of observations. GOF indices summarise the discrepancy between the observed values and the values expected under a statistical model. Likelihood is proportional to the probability of the observed data, so the best-fit parameter value is achieved when the parameter value maximises the likelihood. As likelihoods might be very small numbers, the logarithm of likelihood is used; in this case, the highest values of log-likelihood correspond to the best parameter estimate. With L denoting the log-likelihood, two common GOF indices are Akaike's information criterion (AIC) and Schwarz's Bayesian information criterion (BIC). The AIC and BIC are not used to test the model in the sense of hypothesis testing, but for model selection as in this research, where we want to decide which is the best of a selection of models. AIC provides a 'measure' of the relative quality of a statistical model; the test is based on information theory. When several candidate models can be satisfactorily implemented with the same dataset, the model with the lowest AIC is the 'best' model of all those specified for the data. BIC is closely related to AIC, with the difference that the BIC penalty term is more stringent than that for AIC. As with AIC, the smallest BIC indicates the model that best fits the dataset.

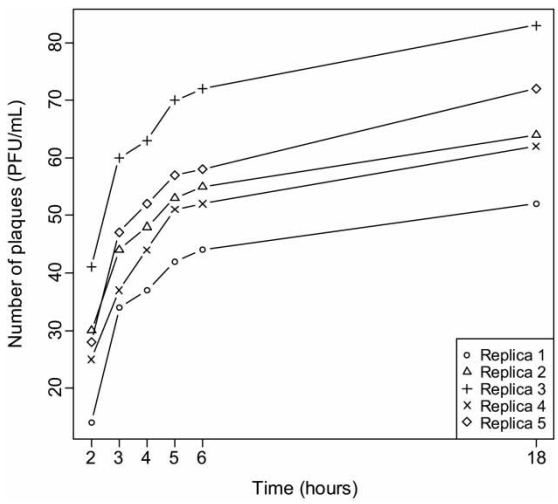

Figure 2 | Plaque counting (replica) in a single river sample coliphage titration.

Briefly, our approach was based on a preliminary test of several regression models using NLME allowing only one parameter of the model to be expressed as a function of a fixed value and a random function. Once the best model was selected, in order to eliminate the need to deal with this variable parameter, a relationship between this parameter and the plate counts after 3 hours was used.

The best fits were using the Antoine expanded or vapour pressure function and the logistic-exponential models. A preference order was established according to their AIC, BIC and determination coefficient (R^2).

Table 3 shows a summary of the equations of some of the models used. The best fits were attained using exponential family curves.

In the modelling, the variables *a*, *b*, *c* and *d*, were taken as fixed components and the random component was related only to the *a* variable (despite randomisation of all variables resulting in a better regression model, this increased the complexity of the predictive model, with the increase in correlation being minimal). Meanwhile, it was only possible to correlate the randomised variable with the value of the phage counts at 3 hours when a stochastic effect was only introduced in one variable.

When NLME was applied separately to each type of sample (Table 1), the logistic-exponential model showed the best fit in several categories; but when NLME was applied to all the samples, the Antoine equation was the most accurate function, resulting in its second position as an alternative to the exponential model; results are shown in Table 3.

Table 3 | Models satisfactorily run using NLME modelling

		Pseudo-R^2 *AIC* Samples		
Model	**Equation**	**Undiluted river**	**Sewage 1:10,000**	**All samples**
Logistic-exponential	$Count\ (PFP) \sim (\frac{a}{1+e^{(b\times Time)}} + c)$	0.979, *2,028.175*	0.884, *423.043*	0.992, *10,408.31*
Antoine expanded	$Count\ (PFP) \sim e^{\left(a+\frac{b}{Time}+c\times\ln(Time)\right)}$	0.978, *2,038.338*	0.882, *420.887*	0.995, *9,712.877*
Gompertz	$Count\ (PFP) \sim a\times e^{-e^{(b-c*\mathrm{Time})}}$	0.975, *2,065.705*	0.882, *425.045*	0.994, *10,065.73*
Monomolecular	$Count\ (PFP) \sim a\times(1-e^{(-b*\mathrm{Time})})$	0.970, *2,105.637*	0.882, *422.698*	0.994, *10,044.56*
Modified Hoerl	$Count\ (PFP) \sim a\times b^{Time}\times Time^{c}$	0.970, *2,106.003*	0.872, *433.002*	0.993, *10,177.52*
Heat capacity	$Count\ (PFP) \sim a + b\times Time + \frac{c}{Time^2}$	0.968, *2,117.821*	0.866, *437.514*	0.972, *11,622.05*
Root	$Count\ (PFP) \sim a\times b^{\left(\frac{1}{Time}\right)}$	0.967, *2,126.023*	0.883, *422.362*	0.993, *10,289.25*
Saturation growth	$Count\ (PFP) \sim \frac{a\times Time}{b+Time}$	0.949, *2,221.43*	0.877, *427.617*	0.990, *10,535.48*
Asymptotic regression	$Count\ (PFP) \sim (a\times e^{(b\times Time)}) + c$	0.828, *2,494.799*	0.828, *463.960*	0.970, *11,702.63*
3rd Polynomic	$Count\ (PFP) \sim (a\times Time) + (b\times Time^2) + (c\times Time^3) + d$	0.752, *2,558.244*	0.666, *527.072*	0.673, *14,024.45*

The NLME model using the Antoine equation can be expressed as shown in Equation (1):

$$Count(PFP) \sim e^{(4.329-(1.962/Time)\ -0.143\times\ln(Time))} \quad (1)$$

The model had an R^2 value of 0.995, with AIC, BIC and log-likelihood values being, respectively, 9,712.877, 9,738.378 and −4,851.439. Estimated values of the variables *a*, *b* and *c* with their respective 95% confidence intervals were 4.329 (4.123 to 4.534); −1.962 (−2.042 to −1.881); and −0.143 (−0.157 to −0.128).

The model was shown to be adequate to fit plaque counting over time, when it was applied with concentrations of bacteriophages of between 0 and 400 PFU/mL.

Figure 3 displays the individual fits using the Antoine equation. Although in the solution, the unique *a*, *b* and *c* values are given for the overall NLME, every individual count has its own regression equation. In our particular case, the variable *a* could be expressed as a function of a fixed value (4.329) and a random function (≈ ±2.2). The variables, *b* and *c* remained fixed with values of −1.962 and −0.143, respectively.

Here, it is important to note that in this model the random effects proved to be highly related to the quantity of the bacteriophage in the water matrices (Spearman's rho of 0.997), this relation being linear-log (R^2 of 0.995) for samples containing more than 10 PFU/mL. However, they showed little association with water sample origins (river or wastewater), not displaying differences ($p > 0.05$, Mann–Whitney U test) between matrices when the phage contents were similar in these samples.

Prediction functions

In order to predict the possible number of coliphages at 18 hours by an early count, two statistical approximations

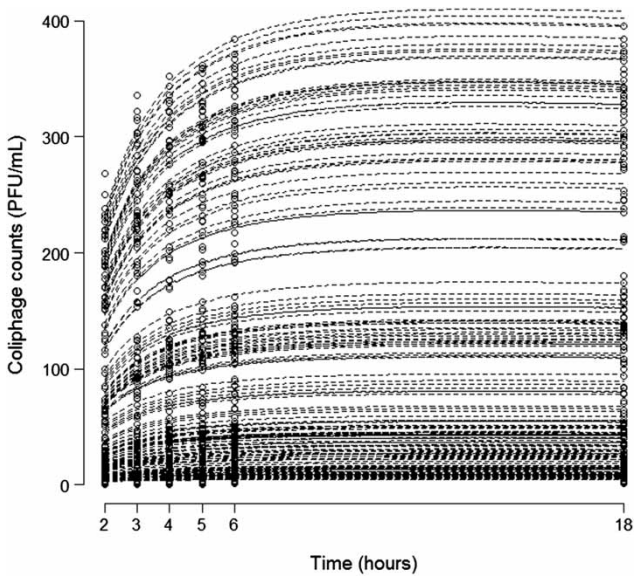

Figure 3 | Individual fits using the Antoine model.

were evaluated. The first one involved the use of the results of the NLME analysis and the second one involved the use of the linear regression and linear mixed-effects (LME) modelling analysis.

All data with the NLME Antoine equation

Commonly, prediction assessment from random-effects models requires a more complex approximation because the solutions involve matrices, and the confidence interval for prediction needs to be calculated from the data used in the modelling.

As was noted before, the origin of the samples has little impact on the model, on the contrary, the variability in the number of bacteriophages was the main factor responsible for the mixed effects. However, it is important to note that the variation in the variable *a* modifies the shape of the predicting curve, and this change allows prediction of the kinetics of the appearance of the plaques for each sample. As the variable *a* was clearly associated with the counts of bacteriophages, we used a numerical approximation in order to establish a relationship between the plaque counts at 3 hours and the variable *a*.

That approximation allows us to predict the coliphage counts at different incubation times (*inc. hours*), as is shown in Equation (2):

$$Counts_{inc.hours} = e^{\left(\left(\frac{2.075+0.235\times Count_{3h}}{1+0.038\times Count_{3h}-1.350\cdot 10^{-5}\times (Count_{3h})^2}\right)-\frac{1.962}{inc.hours}-0.143\times\ln(inc.hours)\right)} \quad (2)$$

Note that Equation (2) allows determination of the number of bacteriophages in any incubation time; the R^2 values for the predicted counts after 4, 5, 6 and 18 hours of incubation were, respectively, 0.997, 0.994, 0.991 and 0.988.

For incubations after 18 hours, as is indicated in the ISO standard, the AIC, BIC and log-likelihood achieved were: 1,620.232, 1,630.156 and −807.1158, respectively, being the 95% confidence interval for fit values of ±2.52 PFU/mL and ±28.18 PFU/mL for predictions.

This approximation is able to determine bacteriophage counts beyond 3 incubation hours, and it showed good linearity between counts at 3 hours and at 18 hours. That linearity is maintained with coliphage values of lower than 350 PFU/mL measured at 3 hours, as shown in Figure 4. Despite this fact, it is necessary to keep in mind that it is recommended to dilute the sample so as to enumerate less than 300 plaques per plate.

Note in Figure 4(a) that in spite of the linearity, the curve has a slight S-shape and the intercept was not located at zero. This denotes that the prediction fails with zero values; a zero value at 3 hours does not suggest that new plaques cannot appear, but neither that a zero value will be obtained at the end of the incubation.

As was indicated previously, a second statistical approximation, using linear regression and LME modelling, was considered in order to simplify calculation, and mainly, for diminishing the prediction interval, that is especially useful in samples containing a low number of bacteriophages. These linear models only have to take into account the bacteriophage counts at 3 hours and they allow prediction of the final counts, as is indicated in the ISO standard, after 18 hours of incubation, but they cannot be applied to predict the kinetic of the appearance of coliphage plaques before that incubation period.

Linear regression and linear mixed modelling

Overall data: Linear regression resulted in an R^2 of 0.9858; and AIC, BIC and log-likelihood of 1,650.165, 1,660.090 and −822.082, respectively. The intercept was 9.526 (95% CI 6.966 to 12.085) and the slope was 1.246 (95% CI 1.225 to 1.267), so $Fit\,Count_{18h} = 1.24611 \times Count_{3h} + 9.52552$. In addition, the 95% confidence interval was ±2.52 PFU/mL for fit values with a prediction interval of ±28.18 PFU/mL.

Meanwhile, LME with random and fixed intercept and slope provided a slightly better fit, with an R^2 of 0.997 and AIC, BIC and log-likelihood values of 1,511.048, 1,530.898 and −749, respectively. The intercept was 14.628 (95% CI 9.493 to 19.763) and the slope was 1.240 (95% CI 1.190 to 1.290). The higher value of the intercept and the similar slope indicate that low concentrations of bacteriophages may be overestimated; consequently, the linear regression model is preferred.

Split data: Due to the value of the intercept, samples containing small amounts of bacteriophages were overestimated. To deal with this fact, the coliphage counts were split into two groups. This increased the accurateness in the prediction

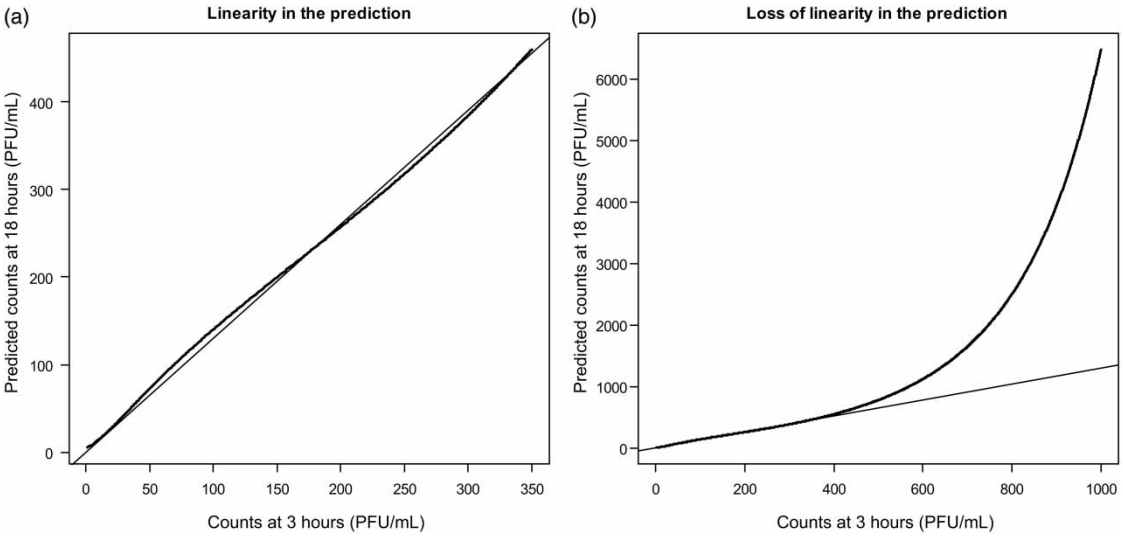

Figure 4 | Linearity between predictors and their predicted values. The thin line represents a slope of 1.3. (a) Linearity below 350 PFU/mL. (b) Shows that the linearity did not extend beyond 350 PFU/mL.

of low values. All possible linear regressions were carried out to find a cut-off value to maximise prediction accuracy, which resulted in 26 PFU/mL at 3 hours of incubation.

Linear regression with samples containing <26 PFU/mL resulted in an R^2 of 0.900; while AIC, BIC and log-likelihood values were, respectively, 496.753, 504.319 and −245.377. The intercept was 1.966 (95% CI 0.797 to 3.135) and the slope 1.480 (95% CI 1.377 to 1.582); the confidence interval at 95% for fit values was ±0.95 PFU/mL and the resultant prediction interval was ±7.06 PFU/mL. Mixed linear models with a random intercept and slope resulted in an R^2 of 0.949 and AIC, BIC and log-likelihood values were: 481.328, 496.459, and −234.664, respectively; the intercept was 2.370 (95% CI 1.074 to 3.667) and the slope 1.181 (95% CI 1.275 to 1.605). Again, linear regression was preferred to predict low coliphage concentrations.

Linear regression with samples containing ≥26 PFU/mL resulted in an R^2 of 0.981; and AIC, BIC and log-likelihood values were: 934.564, 942.666 and −464.282, respectively. The intercept was 22.952 (95% CI 17.481 to 28.423) and the slope 1.181 (95% CI 1.148 to 1.215); confidence interval at 95% for fit values of ±4.29 PFU/mL with a prediction interval of ±33.23 PFU/mL. The mixed linear models with random intercept and slope resulted in an R^2 of 0.995 and AIC, BIC and log-likelihood values were 882.996, 899.199 and −435.498, respectively. The intercept was 24.392 (95% CI 15.951 to 32.834) and the slope 1.190 (95% CI 1.129 to 1.251).

Although LME models provided better fits than linear regression, here, both models showed similar results for slope and intercept. Thus, as both models may be used indistinctly, and because the prediction interval is easier to calculate with linear regression, the LME model could be discarded. Figure 5 shows the scatterplot of the 95% confidence and prediction interval of the regressions.

Limits of use of early counting

The limit of use of early counting was determined using a regression analysis with 10,000-fold diluted sewage water samples, which included zero values. The value of sensitivity obtained was 3.64 PFU/mL, calculated from the prediction interval at 95%; consequently, when coliphage counts are ≤4, it is recommended to extend the incubation time to (18 ± 3) hours, as indicated in the ISO standard.

DISCUSSION

Several techniques have been reported to shorten the time required to obtain somatic coliphage presence/absence results.

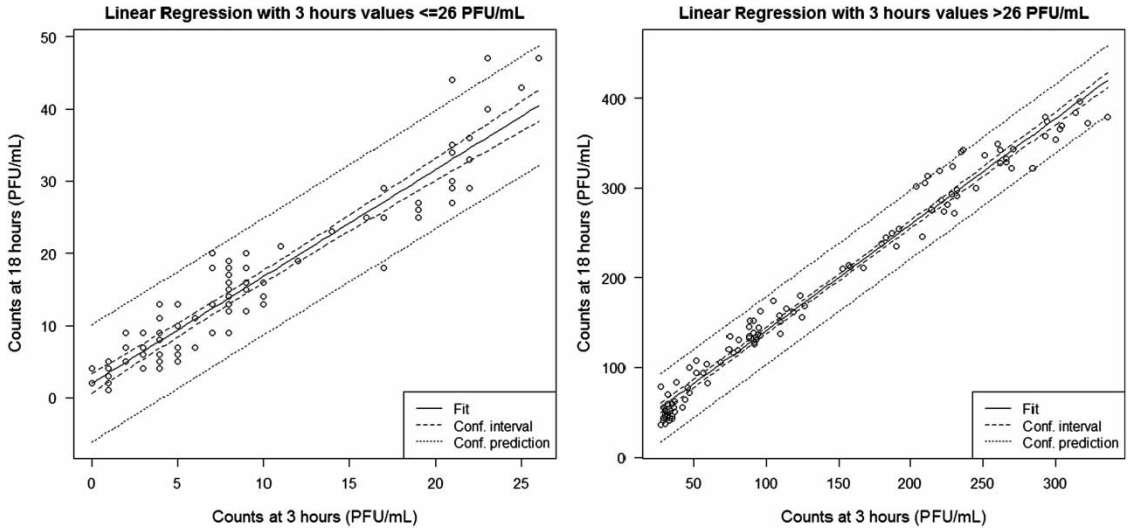

Figure 5 | Results of lineal regression below and beyond 26 plaques by plate at 3 hours.

Armon & Kott (1993) developed a rapid presence detection method for 500 mL drinking water samples using a jar containing the water sample and an immersible probe filled with solidified soft agar containing *E. coli* CN13 as the bacterial host that detected somatic coliphages within 3 hours. Kim *et al.* (2009) developed a bioluminescence-based assay for enumeration of lytic bacteriophages that uses transformed *E. coli* DH5α with a pCRII cloning vector containing a luxCDABE gene cassette. That method allows the detection of high levels of T4 phage (8.4 log PFU/mL) within 1.3 hours. Guzmán Luna *et al.* (2009) developed an enumeration method based on detection via a bioluminescent assay of the release of the bacterial host adenosine 5′-triphosphate and adenylate kinase, due to phage-mediated lysis. That method provides results in less than 3 hours for >10 PFU using *E. coli* WG5 as the bacterial host. More recently, Wang & Nitin (2014) developed a rapid detection of bacteriophage method in starter cultures based on water-in-oil-in-water emulsion micro-droplets that allows the detection of bacteriophages in a minimal time of 1 hour with a detection limit of 100 PFU/mL. Nevertheless, little attention has been devoted to establishing the minimum incubation time of traditional methods to provide a reliable result.

In this article, different prediction models are shown. Both linear models were successfully applied to predict the bacteriophage counts, whereas the model based on the Antoine equation was also able to predict the plaque appearance kinetics. The prediction achievements of the three models were similar, with R^2 ranging from 0.981 to 0.995; but the main difference between them arose with the prediction intervals. However, in the case of samples containing more than 26 PFU/mL the prediction intervals between the linear model and the model based on the Antoine were similar. In samples containing less than 26 PFU/mL, the prediction interval of the linear model was much lower, which indicates that this model should be preferred for samples containing a low number of bacteriophages. Thus, linear regressive models can be used as long as the kinetic need not be evaluated; in that case, the model based on the Antoine equation should be applied.

The obtained regressive models reduce the detection limit to 4 PFUs and allow us to obtain a coliphage estimation after 3 hours of incubation. Meanwhile, variable overlap between plaques has been observed, so the single count may tend to diminish the real number of bacteriophages detected; therefore it would be advisable to perform at least one additional count within the first 6 hours of incubation, especially in cases where the counts have to fulfil a compulsory or recommended threshold value.

CONCLUSIONS

Early plate counts offer a feasible and reliable procedure for the detection of phages at levels greater than 4 PFU/mL of somatic coliphages in water samples after 3 hours of incubation, using the standard coliphage quantification procedure.

Considering the amounts of bacteriophages described in several types of waters and their suggested quality criteria, this early count could be applied in water to monitor the efficiency of wastewater and water reclamation treatments, and to indicate the quality of bathing waters and reclaimed waters with quantifiable results on the same working day.

The results reported here indicate that the traditional coliphage plating technique has the potential for very rapid, practical, real-time detection that can be implemented without inconvenience by laboratories that routinely titrate bacteriophages.

ACKNOWLEDGEMENTS

This study was supported by the Spanish Ministry of Education and Science (CGL2011-25401), the FP7 KBBE project (AQUAVALENS), the *Generalitat de Catalunya* regional authorities (2014 SGR0007) and by the Catalan Reference Network of Biotechnology (XRB). No conflict of interest is declared.

REFERENCES

Anonymous 2001 Loi sur la qualité de l'environnement: règlement sur la qualité de l'eau potable c. Q-2, r.18.1.1. [Environment Quality Act: Drinking Water Quality Regulation c. Q-2. R-18.11]. *Gazette Officielle du Québec*, **24**, 3561. Government of Quebec, Montreal, Quebec, Canada.

Anonymous 2005 *Queensland Water Recycling Guidelines.* State of Queensland, Brisbane, Australia.

APHA-AWWA-WPCF 2001 *Standard Methods for the Examination of Water and Wastewater*, 20th edn. American Public Health Association, Washington, DC.

Armon, R. & Kott, Y. 1993 A simple, rapid and sensitive presence/absence detection test for bacteriophage in drinking water. *J. Appl. Bacteriol.* **74**, 490–496.

Armon, R. & Kott, Y. 1996 Bacteriophages as indicators of pollution. *Crit. Rev. Environ. Sci. Technol.* **26**, 299–335. doi:10.1080/10643389609388494.

Ashbolt, N. J., Grabow, W. O. K. & Snozzi, M. 2001 *Indicators of Microbial Water Quality. Water Quality: Guidelines, Standards and Health* (L. Fewtrell & J. Bartram, eds). IWA Publishing, London, UK.

Bates, D., Maechler, M., Bolker, B. & Walker, S. 2015 Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67** (1), 1–48. doi:10.18637/jss.v067.i01.

Colquhoun, K. O., Timms, S. & Fricker, C. R. 1995 Detection of Escherichia coli in potable water using direct impedance technology. *J. Appl. Bacteriol.* **79**, 635–639.

Gassilloud, B., Schwartzbrod, L. & Gantzer, C. 2003 Presence of viral genomes in mineral water: a sufficient condition to assume infectious risk? *Appl. Environ. Microbiol.* **69**, 3965–3969. doi:10.1128/AEM.69.7.3965-3969.2003.

Grabow, W. O. K. 2001 Bacteriophages: update on application as models for viruses in water. *Water SA* **27**, 251–268. doi:10.4314/wsa.v27i2.4999.

Grothendieck, G. 2013 nls2: Non-linear regression with brute force. R package version 0.2. http://CRAN.R-project.org/package=nls2.

Guzmán, C., Mocé-Llivina, L., Lucena, F. & Jofre, J. 2008 Evaluation of Escherichia coli host strain CB390 for simultaneous detection of somatic and F-specific coliphages. *Appl. Environ. Microbiol.* **74**, 531–534. doi:10.1128/AEM.01710-07.

Guzmán Luna, C., Costán-Longares, A., Lucena, F. & Jofre, J. 2009 Detection of somatic coliphages through a bioluminescence assay measuring phage mediated release of adenylate kinase and adenosine 5′-triphosphate. *J. Virol. Methods* **161**, 107–113. doi:10.1016/j.jviromet.2009.05.021.

Hsu, F. C., Shieh, Y. S. C. & Sobsey, M. D. 2002 Enteric bacteriophages as potential fecal indicators in ground beef and poultry meat. *J. Food Prot.* **65**, 93–99.

IAWPRC Study Group on Health Related Water Microbiology 1991 Bacteriophages as model viruses in water quality control. *Water Res.* **25**, 529–545.

International Organization for Standardization 1995 *Water Quality: Detection and Enumeration of Bacteriophages. 2. Enumeration of Somatic Coliphages*. ISO 10705-2. International Organization for Standardization, Geneva, Switzerland.

Jofre, J. & Blanch, A. 2010 Feasibility of methods based on nucleic acid amplification techniques to fulfil the requirements for microbiological analysis of water quality. *J. Appl. Microbiol.* **109**, 1853–1867. doi: 10.1111/j.1365-2672.2010.04830.x.

Kim, S., Schuler, B., Terekhov, A., Auer, J., Mauer, L. J., Perry, L. & Applegate, B. 2009 A bioluminescence-based assay for enumeration of lytic bacteriophage. *J. Microbiol. Methods* **79**, 18–22. doi:10.1016/j.mimet.2009.07.011.

Kott, Y., Roze, N., Sperber, S. & Betzer, N. 1974 Bacteriophages as viral pollution indicators. *Water Res.* **8** (3), 165–171. doi:10.1016/0043-1354(74)90039-6.

Loader, C. 2013 locfit: Local Regression, Likelihood and Density Estimation. R package version 1.5-9.1. http://CRAN.R-project.org/package=locfit.

Lucena, F. & Jofre, J. 2010 Potential use of bacteriophages as indicators of water quality and wastewater treatment processes. In: *Bacteriophages in the Control of Food- and Waterborne Pathogens* (P. M. Sabour & M. W. Griffiths, eds). ASM Press, Washington, DC, pp. 103–118.

Mendez, J., Jofre, J., Lucena, F., Contreras, N., Mooijman, K. & Araujo, R. 2002 Conservation of phage reference materials and water samples containing bacteriophages of enteric bacteria. *J. Virol. Methods* **106**, 215–224. doi:10.1016/S0166-0934(02)00163-5.

Park, S. J., Lee, E. J., Lee, D. H., Lee, S. H. & Kim, S. J. 1995 Spectrofluorometric assay for rapid detection of total and fecal coliforms from surface water. *Appl. Environ. Microbiol.* **61**, 2027–2029.

Pillai, S. D. 2006 Bacteriophages as fecal indicator organisms. In: *Viruses in Foods* (S. M. Goyal, ed.). Springer, New York, NY.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. & R Core Team 2015 nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-120. http://CRAN.R-project.org/package=nlme.

Rasband, W. S. & ImageJ, U. S. 1997–2016 National Institutes of Health, Bethesda, MD, USA. http://imagej.nih.gov/ij/.

R Development Core Team 2014 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Reasoner, D. J. & Geldreich, E. E. 1989 Detection of fecal coliforms in water by using [14C]mannitol. *Appl. Environ. Microbiol.* **55**, 907–911.

Sobsey, M. D., Battigelli, D. A., Shin, G. A. & Newland, S. 1998 RT-PCR amplification detects inactivated viruses in water and wastewater. *Water Sci. Technol.* **38** (12), 91–94. doi:10.1016/S0273-1223(98)00807-5.

US Environmental Protection Agency 2001a *Method 1601. Male-Specific (F+) and Somatic Coliphage in Water by Two-Step Enrichment Procedure*. EPA report no. 821-R-01-030. Office of Water, Engineering and Analysis Division, Washington, DC.

US Environmental Protection Agency 2001b *Method 1602. Male-Specific (F+) and Somatic Coliphage in Water by Single Agar Layer Procedure*. EPA report no. 821-R-01-029. Office of Water, Engineering and Analysis Division, Washington, DC.

US Environmental Protection Agency 2006 *National Primary Drinking Water Regulations: Groundwater Rule*. Final Rule; 40 CFR Parts 9, 141 and 142. Federal register, Vol. 71, No. 216. Office of the Federal Register, National Archives and Records Service, General Services, Washington, DC, pp. 65574–65660. Available from https://www.federalregister.gov/documents/2006/11/08/06-8763/national-primary-drinking-water-regulations-ground-water-rule.

US Environmental Protection Agency 2012 *Guidelines for Water Reuse*. EPA/600/R-12/618. US EPA Office of Research and Development, Washington, DC. Available from http://nepis.epa.gov/Adobe/PDF/P100FS7K.pdf.

Venter, S. N., Catalan, V., du Preez, M., Everaert, J., Finch, P., Fricker, C., Grant, M., Levine, A. & Seemendi, S. 2000 *Rapid Microbiological Monitoring Methods: The Status Quo*. International Water Association's Blue Pages. IWA, London, UK.

Wang, M. S. & Nitin, N. 2014 Rapid detection of bacteriophages in starter culture using water-in-oil-in-water emulsion microdroplets. *Appl. Microbiol. Biotechnol.* **98** (19), 8347–8355. doi: 10.1007/s00253-014-6018-7.

World Health Organization 2008 *The Global Burden of Disease: 2004 Update*. WHO, Geneva, Switzerland. doi:10.1038/npp.2011.85.

First received 27 February 2017; accepted in revised form 29 May 2017. Available online 13 July 2017