Appendix 2: Validation and Quantification of phytoplankton using qPCR

Phytoplankton growth and cell counts

To convert DNA copy numbers into cell numbers. Cultured representatives of Phytoplankton being tested were grown and counted. DNA from *P. fraudulenta* (either from CCAP1061/4 or CCAP 1061/6), deposited by J. Fehling in 2001 *Pseudo-nitzschia fraudulenta* cell culture IO83-07, University of Lisboa, Portugal and culture CCAP1061/4 or CCAP 1061/6 (deposited by J. Fehling in 2001) was grown in f/2 media (Guillard 1962, Guillard 1975) at 15ºC, with a 14:10 light:dark cycle. For cell counts, cell culture IO83-07 were grown to exponential phase and the sample collected 5 days after inoculation. Two replicate samples (2ml) were collected into sterile RNAase free microfuge tubes and two replicate sets were prepared as a serial dilution (1:10, 1:100, 1:1000 and 1:10000) by successive transfers of 0.2 ml of one concentration to a new eppendorf containing 1.8 ml of sterile f/2 medium. The samples were then centrifuged at 3000g (4000 rpm, Labnet Spectrafuge 16M microcentrifuge) for 5 minutes. The supernatant was pipetted off and a final volume of 0.1 ml was left with the cell pellet to reduce possibility of removing cells. RNAlater was then added to the sample for storage at 4°C. An aliquot of the culture was well mixed with an automatic pipette to homogenizing the sample and break the cell chains to guarantee a better estimation of the cell concentration. The flask was also rotated between sub-sampling. Cells were counted in a Neubauer counting chamber. A total of 313 cells were counted giving an estimated error around the mean of 11% giving a live cell count of 17389 ± 1966 cells.ml^{-1} and total cell count (live cells + dead cells) of 18833 ± 2046 cells.ml^{-1}. The dead cell count was used for calculating ITS copies within *P. fraudulenta*. These counts were used to convert DNA copy numbers to cell numbers. No counts were made of *P. fraudulenta* CCAP 1061/4 or 6. DNA from both cultures were run side by side alongside cloned standards and they were found to have similar profiles (see below) so counts of IO83-07 were used for all cell counts

Pseudo-nitzschia delicatissima (CCAP 1061/41) culture was stored in a 15°C temperature controlled room and kept in F/2 media (Guillard and Ryther 1962; Guillard 1975). Prior to DNA extraction, 1ml was removed and lugols was added for preservation. This sample was used for cell counts using a Sedgewick Rafter counter under a light microscope (Optech microscopes LTD) at x20 magnification at the same time as the culture was used for DNA extraction.

To calculate the cell density of the *Aureococcus anophagefferens* cells (RCC4094) culture, 50 squares (50µl) of the Sedgewick rafter was counted and averaged. A 1ml aliquot of *Aureococcus anophagefferens* cells (RCC4094) were removed from culture and counted using a Sedgewick Rafter counter under a light microscope (Optech microscopes LTD) at x20 and x40 magnification at the same time as the culture was used for DNA extraction.

DNA extraction

DNA from *P. fraudulenta* (either from CCAP1061/4 or CCAP 1061/6, deposited by J. Fehling in 2001 and the culture now died) was extracted using Qiagen DNeasy blood and Tissue kit (Qiagen, UK). *Pseudo-nitzschia delicatissima* (CCAP 1061/41) culture was extracted using CTAB method followed by phenol-chloroform extraction and ethanol precipitation as described by Clarke et al. (2015). The DNA was quantified using quantifluor (Promega, Madison, USA) according to the manufacturers instructions. *Aureococcus anophagefferens* cells were extracted using Qiagen DNeasy blood and Tissue kit (Qiagen, UK) according to manufacturers instructions.

Preparation of HRM-qPCR Standards

PCR products were amplified using primers in Table 2 using GoTaq PCR kit (Promega). *P. fraudulenta* ITS amplicons were generated from the following thermocycling conditions at 95°C for 5 minutes followed by denaturation at 95°C for 30 seconds, annealing at 59°C for 30 second and extension 72°C for 30 seconds for 35 cycles, with a final extension at 72°C for 7 minutes. *P. delicatissima* ITS amplicons were amplified using the same conditions except denaturation was at 60°C and the final extension time was for 5 minutes. The 18S qPCR assay developed by Popels et al. (2003), with corresponding species primers in Table 2 using Popels et al. (2003) thermocycling conditions except using JumpStart Redtaq ReadyMix Reaction mix according to the manufacturers instructions (Sigma, now Merck). The ITS standard for Pseudo-nitzschia multiseries was a synthesized ITS marker corresponding to a *P. multiseries* ITS region (201bp in length, position 188-390 derived from sequence *Pseudo-nitzschia multiseries* isolate Pn_1, DQ445651) contained within a pEX-A2 vector produced by MWG-Eurofins (Germany) called Pmult_ITS_pExA2.

PCR products were cloned into the PCR 2.1 vector using TA Cloning Kit with and One Shot TOP10 Chemically Competent *E. coli* (Life Technologies, Paisley, UK). Ligation reactions were completed overnight at 15°C (T4 DNA Ligase) or within 1 hour using ExpressLink™ T4 DNA Ligase (Invitrogen, now Thermofisher, UK) according to manufacturer's instructions and plated onto LB-agar plates with either 100μg/ml ampicillin or 50μ/ml kanamycin, supplemented with 40μl of X-Gal to differentiate between transformants according to manufacturer's instructions. Three colonies were picked and grown in 5 ml of LB media (Bertani, 1951) with 50μg/ml kanamycin overnight. DNA was extracted using DNeasy plasmid mini prep (Qiagen, UK). The cloned insert was sequenced using M13F primers to confirm their identity. The concentration of cloned standards at different dilutions plus genomic DNA from the same culture was established using the QuantiFluor instrument (Promega, Madison, USA), according to manufacturer's instructions.

HRM-qRT-PCR assays

HRM- qRT-PCR assays can both specifically detect species present using High Resolution Melt (HRM) curve of the PCR product and quantify that product using real time PCR simultaneously in the same assay. The concentration of both plasmid or PCR product standards and/or positive control DNA was previously established using the QuantiFluor (Promega, Madison, USA) according to manufacturer's instructions. DNA standards were generated from PCR products of known cultured species representatives or cloned synthesized DNA (for *P. multiseries*) serially diluted to known copy numbers in duplicate or triplicate as described in Appendix 2. Additionally positive (DNA of identified cultured species) was included to assess any melt-curve differences between cloned and genomic DNA and negative (no DNA template) control reactions assured PCR reactions were not contaminated. WaMS DNA samples were diluted 1:10. The standard dilution series was run in triplicate for *A.anophagefferens* qRT-PCR runs AA2011, AA2012. All *P. delicatissima* standards were run in duplicate for statistical accuracy.

Each reaction contained 5μL of SensiFAST HRM Kit (Bioline Reagents Limited, London) 0.4μl of each primers, 2μl of 1:10 diluted DNA in a total reaction volume of 10 μl. The optimised thermal cycling conditions were as follows according to their protocol: 1 cycle of 95°C for 3min for polymerase activation, followed by 40 cycles of denaturation at 95°C for 5s, primer annealing at 60°C for 10s and extension at 72°C 20s. Following the cycling phase, HRM melting phase ranged from 70-90 °C, rising in 0.1 °C increments.

3.2.1 Validation of Pseudo-nitzschia assay HRM curve specificity and qRT-PCR reaction parameters

HRM curves from plasmid standards and genomic DNA from positive controls revealed a specific peak at 84-84.52°C for *P. fraudulenta* (strain PLY1St.36A), with 84°C observed for standards at the lowest concentration, although non-specific peaks were identified outside its range which may be the reason for reduced reaction efficiency. Sequencing the ITS2 region of the genomic DNA control confirmed it was *P. fraudulenta* (LN873237). Sequencing *P. fraudulenta* ITS2 amplicons after qRT-PCR from WaMS DNA in samples E4_5.1/WS12 confirmed the product was ITS2 from *P. fraudulenta* (LN873238). *P. multiseries* HRM curve showed a specific peak range of 82.47-82.7. The qRT-PCR derived amplicons from the positive *P. multiseries* genomic control (Pm12) confirmed its identity (accession number OP504083) No significant differences in HRM curve range was found between 2011-2012 samples for *P. multiseries* qRT-PCR however some non-specific amplicons were also observed. Sequencing of positive control `qRT-PCR` amplicons confirmed the assay detected *P. delicatissima* ITS1 (OM350397).

The plasmid standard curves for *P. fraudulenta* and *P. multiseries* had good correlation coefficient value ($R^2>0.99$) and consistency between standard Ct value between replicates but their efficiency figures were lower than the recommended 90% on the standard curve ranging from 70-72% for *P. fraudulenta* and 66-70% for *P. multiseries* likely reflected in lowered overall reaction efficiency values of the standard curve slope (M: -4.236 to -4.56) which should be near -3.322. Due to their lowered efficiency the results were not deemed to be reliable for subsequent interpretation of species cell numbers. HRM analysis of *P. delicatissima* revealed a species-specific peak between 82-82.5C (Figure A1, panel B). *P. delicatissima* reaction efficiency (see Table 3) was between 80-120%, (Figure A2 as example standard curves for qRT-PCR runs of 2012 and 2013 samples). The qRT-PCR run for 2011 samples was 87%, slightly lower than ideal 90% or more efficiency, but a 2-tailed student's t-test (unequal variance assumption) of standard Ct values of 2011 with those from 2012 and 2013 assays revealed no significant difference (p=0.153) at 5% level. Detailed descriptions of *P. fraudulenta*/*P. multiseries* assays are in [1] and for *P. delicatissima* [2].

3.2.2 *Aureococcus anophagefferens* assay HRM and inter-run reaction validation

The PCR product of *Aureococcus* assay developed by Popel et al. [3] had a HRM melt curve peak at 82°C (see Figure A1m panel C) and was specific to *Aureococcus* upon sequencing products from environmental samples E4_10_11 (1)/WS32, E1_12_11/WS34 and E1_2_12 or WS39 (accession numbers OP484337-OP484339 respectively). Runs were named after species initials (AA) and the year of the WaMS samples in the assay. The reaction efficiencies were 0.93 for AA2011 run, 0.82 for AA2012 run and 0.78 and 0.61 for AA2013 part 1 and part 2 respectively. Reactions AA2013 were excluded because of low reaction efficiency for the latter and failed student's t-test for range of standard Ct values in the former case. Standard Ct values for Run AA2012 was not statistically different from those of run AA2012 using a student-s t-test (two sided, unequal variance) p=0.392 at 5% level so were included. Other parameters (M, R, R^2) between runs AA2011 and AA2012 were similar to each other and were within recommended manufacturers guidelines (Qiagen, UK). Run AA2011 had very similar reaction parameters to those published [3]. The potential accuracy of mean detection capability is 1.4 cells on average [3].
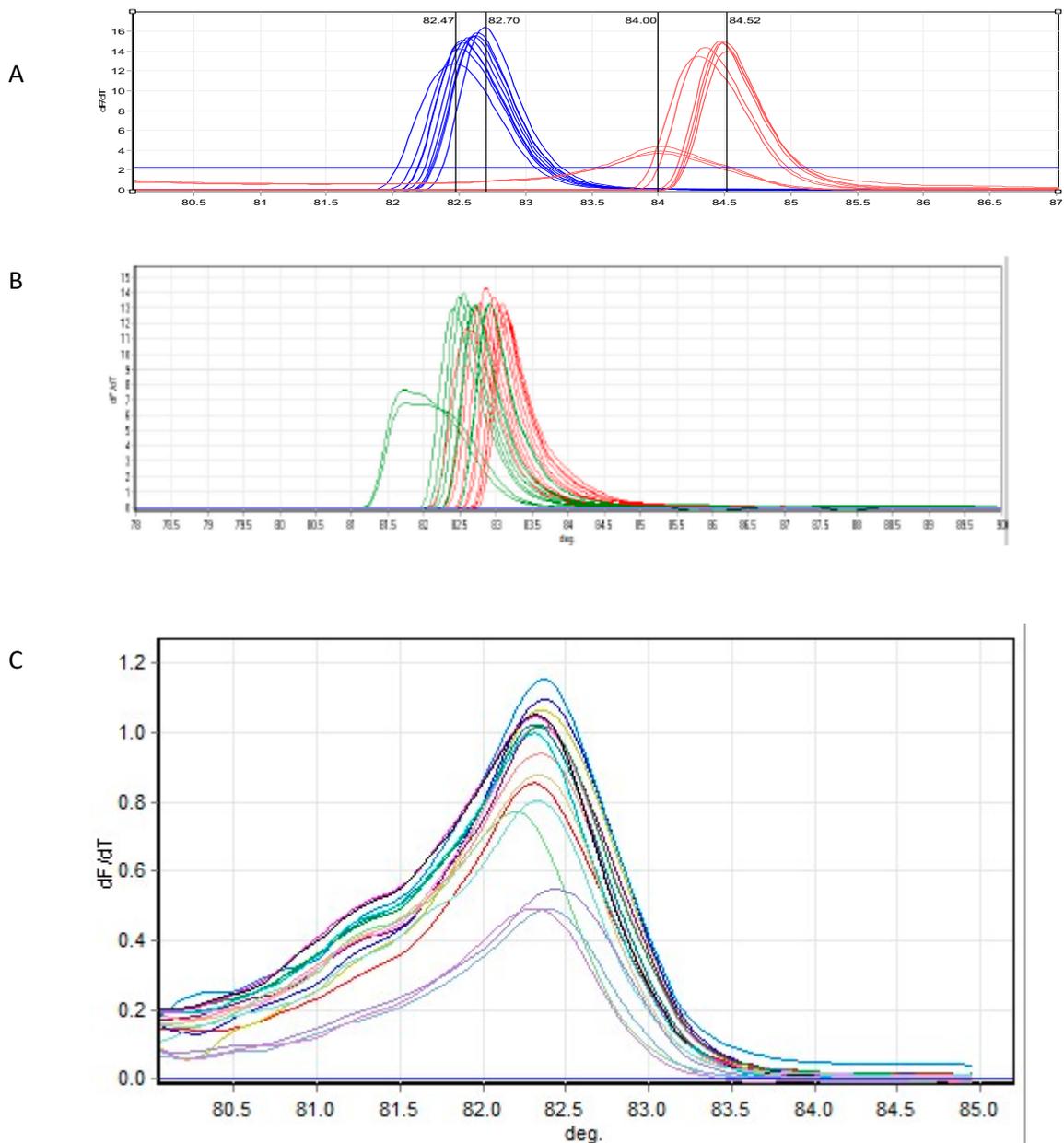
Figure A1. A: HRM melt curves for HRM-QPCR assays for four phytoplankton species. Panel A: *P. multiseries* (blue) and *P. fraudulenta* (red) completed on the same HRM profile. Threshold, set above any noise, indicated by horizontal blue line. Positive controls for *P. fraudulenta* that were more highly diluted (10⁵ to 10⁷) produced the shorter red peaks that were shifted by 0.5℃. Both reproduced from [1]. B: HRM melt curve *P.delicatissima* standards (red) and positive control (green) samples, the genomic DNA sample at 2.5 x 10⁷ showed a shift by 1.5℃. C, reproduced from [2]. *Aureococcus anophagefferens* standards (multiple colours). X axis = Melting temperature. Y axis = the negative derivative (dF/dT) of fluorescence (F) versus temperature (T).
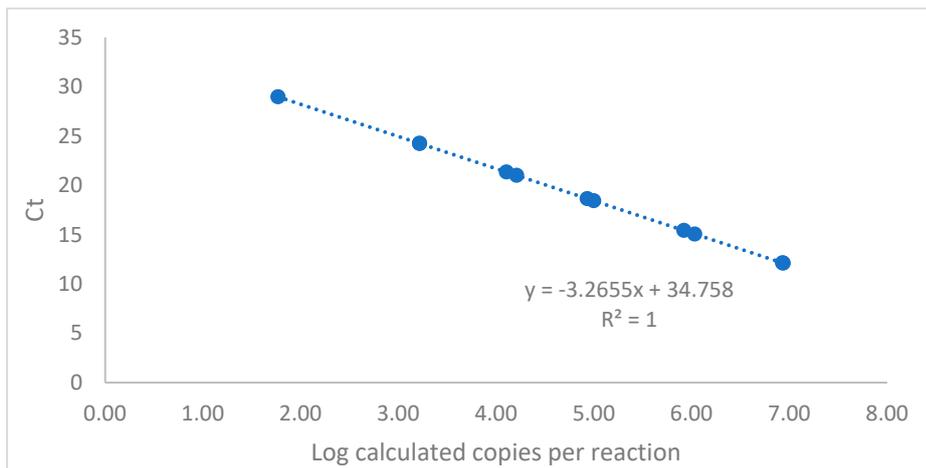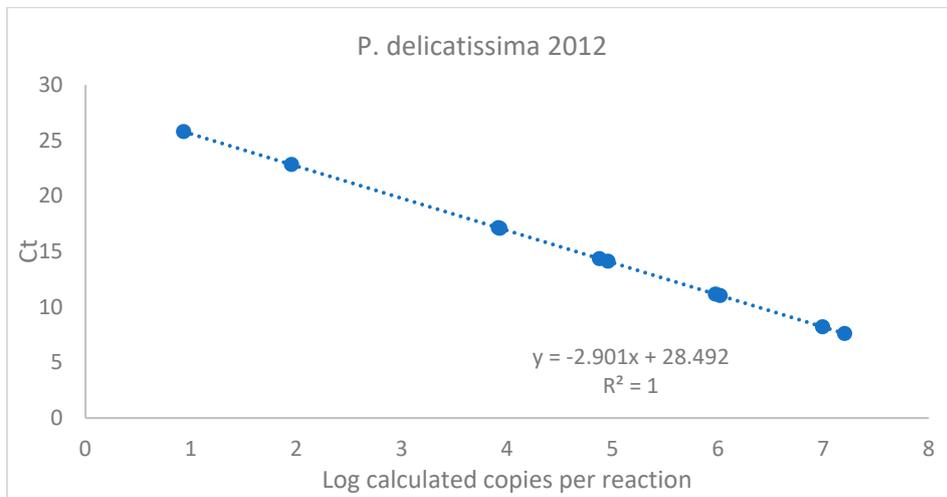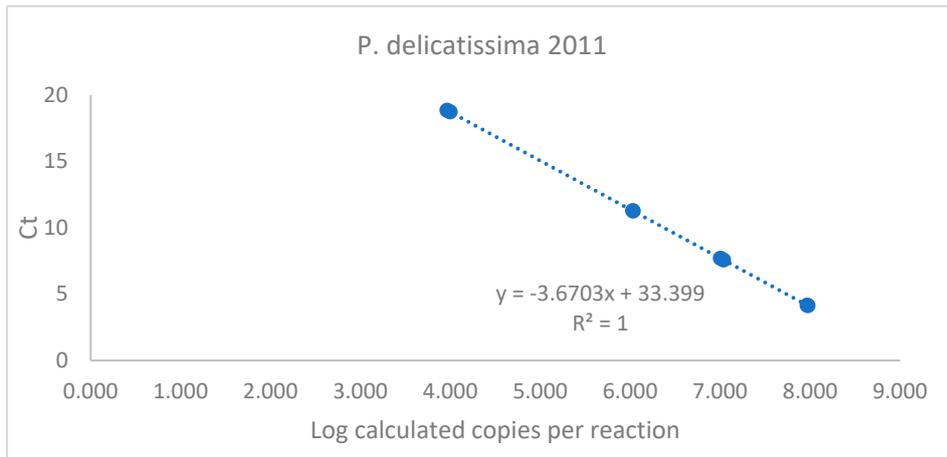
Figure A2. *P. delicatissima* qPCR ITS assay standard curve for 2011, 2012 and 2013 (from top to bottom).

Table A1. Reaction efficiency of qPCR assays in this study PF (*P. fraudulenta)*, PM (*P. multiseries*) for assay run 2011 and 2012. E=efficiency, M=slope, B=offset, R/ R²= correlation coefficient. Efficiencies were below 80% so these assay results were not analysed further here. M values should be near -3.4.

	PF 2011	PF 2012	PM 2011	PM 2012
R	0.995	0.996	0.999	0.999
R^2	0.991	0.993	0.998	0.999
M	-4.236	-4.352	-4.36	-4.56
B	64.883	58.704	50.189	51.245
E	0.72	0.70	0.70	0.66

Quantification: HRM-qPCR data treatment

Copy numbers were calculated for each of the standards using this Equation [1]: Target abundance = 6.023×10^{23} (copies mol^{-1}) × Standard Concentration (g μl^{-1})/Molecular Weight (g mol^{-1}). Molecular weight was calculated using the average $650 gmol^{-1}$ per base pair.

DNA quantities were converted to copy numbers using this Equation [1]: Target abundance = 6.023×10^{23} (copies mol^{-1}) × Standard Concentration (g μl^{-1})/Molecular Weight (g mol^{-1}). Molecular weight was calculated using the average $650 gmol^{-1}$ per base pair.

Before processing, data was assessed for contamination from PCR products using water controls and discarded if contamination was present. Copy numbers of target amplicons for species were inputted directly into the software before a run, or were converted to copy numbers after the Quantitative real time PCR run. The threshold was automatically set by the software as were reaction parameters, standard curves and values. The software used the following calculations to generate reaction parameters and standard curves.

DNA copy number calculation was carried out:

[2] Y=Mx+B

where Y=Ct, M= slope of the line, x= DNA copy number, B= offset or Y intercept

Reaction efficiency was automatically calculated using the software's formula and correlation coefficient as below

[3] Exponential amplification=$10^{[-1/M]}$

[4] Reaction Efficiency =$[10^{[-1/M]} -1]^2$

P.delicatissima **qRT-PCR calculation of cells numbers**

To determine target gene copy numbers per cell for *P. delicatissima*, the quantitative real time PCR run from 2011 was chosen to calculate numbers as it had most complete set of readings for cell number v copy number. A linear regression line was generated as described above using log cell number and plotted against corresponding log calculated cell number/ml and values taken only within accurate working range based on standards. Target copy numbers per cell was calculated from this and found to be between 15-21 copies/cell. This was similar to average copies/cell reported by other studies for this species: 16 and 36

copies/cell [4, 5]. As 2011 efficiency was lower for runs that analysed 2012, 2013 samples, a student's t-test was performed on 2011 and 2012 Ct values revealing no statistical difference between runs (p=0.075).

[5] Target copies/cell= [1/cell number per reaction] x copy number per reaction

ITS1 copy numbers were converted to cell numbers by dividing the copy number by 21, the highest copy number per cell. As WaMS DNA samples were diluted 1 in 10, this number was multiplied by 10. These figures were normalised to account for different DNA volumes of WaMS DNA samples for February-April 2011 compared to other samples to an equivalent of 1ml.

A. anophagefferens qRT-PCR calculation of cells numbers

Four Quantitative real time PCR runs were performed: AA2011 and AA2012 generating standard curves as redrawn in Figures A3 and A4). DNA concentration of *A. anophagefferens* 18S products were converted to copy numbers using equation [1]. Due to poor accuracy of cell counts in this study using a Sedgwick Rafter Counter, we used standard curve of cells per microlitre versus Ct from the original study that used more accurate flow cytometry counts in [3]. Ct versus cell numbers were plotted using values from [3] which generated a standard curve was Y= -3.4x + 25.4, known as equation [6] and correlation coefficient R^2 value of 0.9897. Next, DNA copy numbers from quantitative real-time PCR runs had to be normalised to the scale of the standard curve produced by the study in [3]. The normalised DNA concentration of standards in copies/microlitre from AA2011 were converted to cells/microlitre using equation [6]. The resulting standard curve equation of the form described in equation [2] of normalised DNA concentration against normalised cell concentrations were used to calculate *A. anophagefferens* cell concentrations in unknown environmental samples. These were scaled up to cells/ml. This process was repeated for all the above mentioned quantitative real-time PCR runs for this species.


Figure: 2011 qPCR run of A. anophagefferens Ct versus Log concentration of standards. $y = -3.5016x + 43.483$, $R^2 = 0.9869$
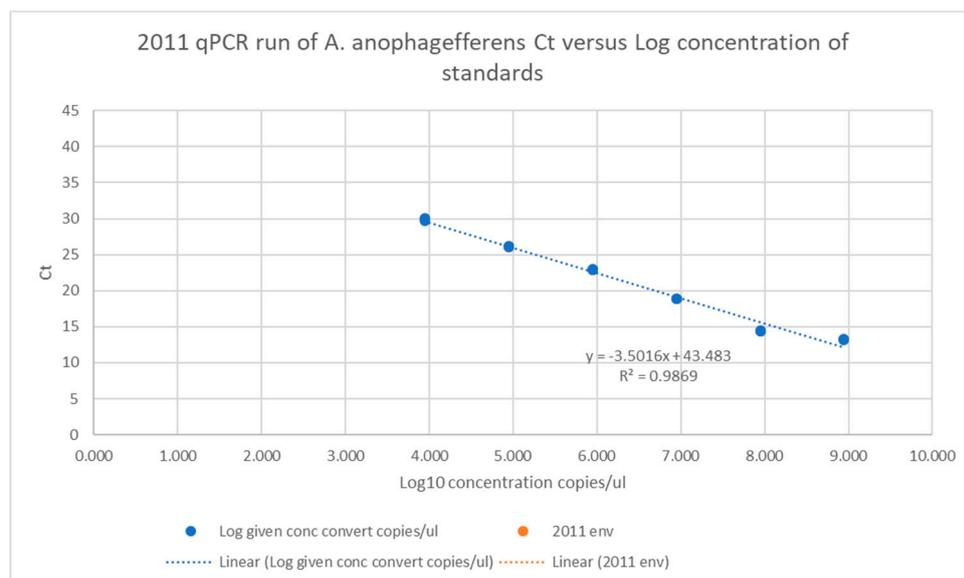
Figure A3. Standard curve of Ct values from 18S quantitative real-time PCR assay of *A. anophagefferens* from AA2011.
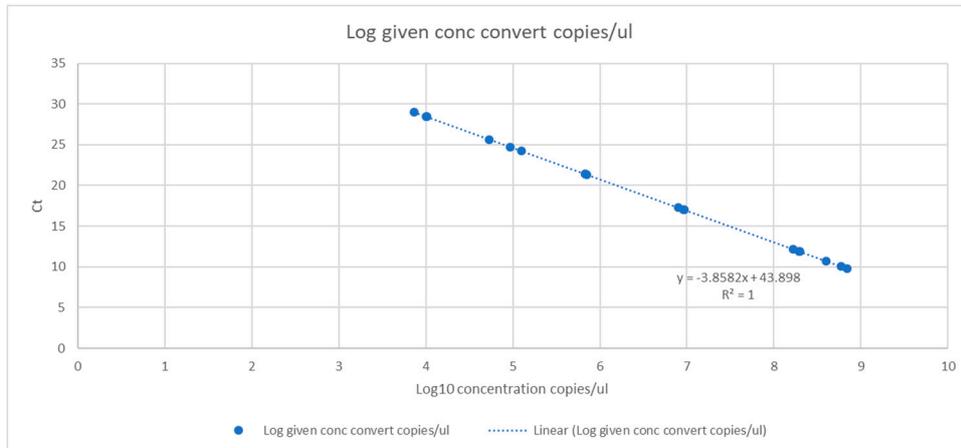

Figure A4. Standard curve of Ct values from 18S quantitative real-time PCR assay of *A. anophagefferens* from AA2012.

References

1.   Walker, C.E., *Molecular identification of Pseudo-nitzschia species in the English Channel.*, in *School of Marine Science and Engineering*. 2014, University of Plymouth: Plymouth. p. 34.

2.   Clarke, J., *Determining the occurence of non-toxic variants of harmful algal species using automated sampling platforms*, in *Faculty of Science and Technology*. 2015, University of Plymouth: Plymouth. p. 73.

3.   Popels, L.C., Cary, S. C., Hutchins, D. A., Forbes, R., Pustizzi, F., Gobler, C. J., Coyne, K. J., *The use of quantitative polymerase chain reaction forthe detection and enumeration of the harmful algaAureococcus anophagef-ferensin environmental samples along the United States East Coast*. Limnol Oceanogr Methods, 2003. **1**: p. 92-102.

4.   Hubbard, K.A., C.H. Olson, and E.V. Armbrust, *Molecular characterization of Pseudo-nitzschia community structure and species ecology in a hydrographically complex estuarine system (Puget Sound, Washington, USA)*. Mar Ecol Prog Ser, 2014. **507**: p. 39-55.

5.   Penna, A., et al., *Toxic Pseudo-nitzschia spp. in the northwestern Adriatic Sea: characterization of species composition by genetic and molecular quantitative analyses*. Journal of Plankton Research, 2012.