# Development and validation of a DNA-based multi-species biomonitoring toolkit using a High-Throughput qPCR platform: a case study of Irish shellfish species

Dennis van der Pouw Kraan[1], Conor Graham[1], Fiona Kavanagh[1], and Luca Mirimin[1]

[1]Atlantic Technological University - Galway City

July 19, 2023

**Abstract**

Biomonitoring of marine life has been enhanced in recent years by the integration of innovative DNA-based approaches, which offer advantages over more laborious conventional techniques (e.g. direct capture) and greater taxonomic resolution especially in complex life cycles and early life stages. However, tradeoffs between throughput, sensitivity and quantitative measurements must be made when choosing between the prevailing molecular methodologies (i.e. metabarcoding or qPCR/dPCR). Thus, the aim of the present study was to demonstrate the utility of a microfluidic-enabled High Throughput quantitative PCR platform (HT-qPCR) for the rapid and cost-effective development and validation of a DNA-based multi-species biomonitoring toolkit, using larvae of 24 commercially targeted bivalve and crustacean species as a case study. The workflow was divided into three main phases: definition of target taxa and establishment of reference databases (PHASE 1); in silico selection/development and in vitro assessment of molecular assays (PHASE 2); and protocol optimization and field validation (PHASE 3). Of a total of 85 assays in silico, 42 were eventually chosen and validated in vitro. Genetic signal showed good correlation with direct visual counts by microscopy, but also showed the ability to provide quantitative data at the highest taxonomic resolution (species level) in a time- and cost-effective fashion. This study developed a biomonitoring toolkit, demonstrating the considerable advantages of this state-of-the-art technology in boosting the development and application of panels of molecular assays for the monitoring and management of natural resources that can be applied to a range of monitoring programmes. Keywords: DNA, High Throughput, qPCR, biomonitoring, shellfish

**Development and validation of a DNA-based multi-species biomonitoring toolkit using a High Throughput qPCR platform: a case study of Irish shellfish species.**

Dennis van der Pouw Kraan, Conor T. Graham, Fiona Kavanagh, Luca Mirimin

Marine and Freshwater Research Centre, Atlantic Technological University (ATU), Dublin Road, Galway, Ireland

Correspondence: dennispouwkraan.research@gmail.com

## ABSTRACT

Biomonitoring of marine life has been enhanced in recent years by the integration of innovative DNA-based approaches, which offer advantages over more laborious conventional techniques (e.g. direct capture) and greater taxonomic resolution especially in complex life cycles and early life stages. However, tradeoffs between throughput, sensitivity and quantitative measurements must be made when choosing between the prevailing molecular methodologies (i.e. metabarcoding or qPCR/dPCR). Thus, the aim of the present study was to demonstrate the utility of a microfluidic-enabled High Throughput quantitative PCR platform (HT-qPCR) for the rapid and cost-effective development and validation of a DNA-based multi-species

biomonitoring toolkit, using larvae of 24 commercially targeted bivalve and crustacean species as a case study. The workflow was divided into three main phases: definition of target taxa and establishment of reference databases (PHASE 1); *in silico* selection/development and *in vitro* assessment of molecular assays (PHASE 2); and protocol optimization and field validation (PHASE 3). Of a total of 85 assays *in silico* , 42 were eventually chosen and validated *in vitro* . Genetic signal showed good correlation with direct visual counts by microscopy, but also showed the ability to provide quantitative data at the highest taxonomic resolution (species level) in a time- and cost-effective fashion. This study developed a biomonitoring toolkit, demonstrating the considerable advantages of this state-of-the-art technology in boosting the development and application of panels of molecular assays for the monitoring and management of natural resources that can be applied to a range of monitoring programmes.

**Keywords** : DNA, High Throughput, qPCR, biomonitoring, shellfish

## 1| INTRODUCTION

Recent advances in molecular technology have boosted the biomonitoring capabilities of DNA-based approaches, particularly in aquatic environments . Such approaches rely on the detection and quantification of DNA copies in an environmental sample, which is used as a method for determining the presence and in some cases the relative abundance or biomass of target species. Since DNA can be present in an environmental sample as both organismal and/or extra-organismal DNA , DNA-based approaches are capable of targeting a wide range of taxa, ranging from micro- to macro-organism (e.g. , encompassing virtually all life-stages from larvae to adults (e.g. . Even though many challenges still exist , the integration of molecular approaches to traditional monitoring techniques such as direct capture, offers an unprecedented opportunity to reduce the cost and increase the throughput of biomonitoring. The development and application of powerful and time/cost-effective biomonitoring tools would benefit many areas of research and applied science, which are usually reliant on labour-intensive screening methods such as microscopic inspection by expert taxonomists .

Generally, DNA-based species detection approaches fall under two main categories: (i) multi-taxa detection, which uses "universal" markers combined with High Throughput Sequencing (HTS) (also known as metabarcoding) and (ii) targeted detection, which typically uses species-specific quantitative real-time PCR (qPCR) assays . Approaches from both categories rely on the availability of robust reference databases such as curated repositories of DNA data from taxonomically identified specimens, ; however while HTS-based methods are more readily available, though require substantial downstream analytical steps, qPCR-based methods require an extensive initial validation phase (i.e. use reference data for assay design), though once established they can be applied rapidly and cost-effectively. The effectiveness of different methods has been compared both in terms of detectability (, as well as biomass of target organisms (, with species-specific qPCR methods showing promising results ().

The initial validation phase required to deem an eDNA qPCR assay reliable can be laborious and demanding in terms of cost and time , however recent advances in real-time qPCR instrumentation have received a boost in integration and throughput capability thanks to the use of microfluidic and nanotechnologies . This includes platforms that can enable qPCR sample/assay Medium- to High-Throughput . Thus, High-Throughout qPCR (HT-qPCR) instruments can play a significant role in speeding up the validation phase of eDNA qPCR assays as well as co-screening samples for multiple targets simultaneously.

For the present study, we chose the monitoring of spawning and larval occurrence/distribution of Irish shellfish species of ecological and commercial importance as a case study. The Irish shellfish industry produces approximately 29,000 tonnes with an estimated value in excess of \euro71 million per annum. The production mainly consists of mussels (*Mytilus spp* ) (\euro17 million, 17.100 tonnes) and oysters (*Magallana gigas, Ostrea edulis* ) (\euro51 million, 11.000 tonnes). Furthermore, many smaller proportions of clam, crab, and lobster species are harvested all around the country (\euro3 million, 900 tonnes) (BIM, 2022). These fisheries are, like in my countries, of national importance as they provide employment in rural maritime communities where employment is especially required to allow people to work and live in these areas. The production of

many shellfish species relies on natural processes of recruitment (i.e., larvae from wild stocks), which can be difficult to predict due to seasonal and interannual environmental variation. Furthermore, shellfish larvae (particularly bivalve species) can be difficult to taxonomically identify during early developmental stages . DNA-based approaches have the potential to overcome such limitations hence improving the biomonitoring capability of screening protocols, as demonstrated by an increasing number of targeted single taxon qPCR studies, including for oysters (*Magallana gigas, Ostrea edulis* ) , mussels (*Mytilus spp* ) , cockles (*Cerastoderma edule* ) , king scallop (*Pecten maximus* ) , razor clams (*Ensis spp.* ) , common clams (*Mya arenaria* ) , and crabs *(Necora puber, Cancer pagurus, Carcinas maenas)* . However, to the best of our knowledge, except for a recent study targeting fish species , High-Throughput multi-species biomonitoring efforts in aquatic ecosystems are limited to using metabarcoding approaches , which often are non-specific and are characterized by uncertain quantitative capability .

Thus, the aim of this study was to use an HT-qPCR platform to describe the development and validation of a multi-species biomonitoring tool to aid the screening of shellfish larvae of ecologic and economic importance in Irish waters.

## 2| MATERIALS AND METHODS

The following workflow was developed by considering prevailing qPCR guidelines , as well as more recent adaptations to environmental DNA screening , while integrating the use of a High-Throughput qPCR platform. As described in detail below, the workflow was subdivided into three main phases consisting of the definition of target taxa and establishment of reference databases (PHASE 1); *in silico*selection and *in vitro* assessment of molecular assays (PHASE 2); and field testing and validation (PHASE 3) (**Figure 1** ).

### 2.1. PHASE 1 - DEFINITION OF TARGET TAXA AND ESTABLISHMENT OF REFERENCE DATABASES

The first step was to identify the main purpose of the toolkit and the research questions being addressed. In the case of this study, the goal was to develop a screening tool capable of determining the presence of bivalve and crustacean larvae in zooplankton samples, with particular interest in the Irish shellfish fisheries and aquaculture industry. Thus, a total of 24 species were chosen based on their ecological and economic importance in the context of Irish coastal waters, including seventeen bivalve species and decapod species (**Table 1** ). To identify the most suitable genes for molecular species identification, a literature review was conducted for previously published studies including genetic markers designed for taxonomic resolution of selected taxa. In order to identify the final target and off-target species occurring in the geographic region of interest (Irish marine coastal waters), a review of organisms of interest and other closely related species (i.e. same Order) was carried out by consulting the World Register of Marine Species (WoRMS) and Ireland's National Biodiversity Data Centre databases (August 2022).

A dedicated reference DNA database was created by combining publicly available barcode sequences with new data from locally sourced organisms. Firstly, based on results obtained from the above review of target and off-target species, previously published DNA sequence data of selected genes from all relevant taxa was obtained from the Barcode Of Life Database (BOLD) and GenBank database. Secondly, to fill any data gaps and confirm target DNA sequences of local taxa, DNA barcode sequence data was also obtained from taxonomically identified voucher specimens (e.g. adult individuals obtained from local markets or opportunistically). For this latter step, DNA was isolated from 50 mg of muscle tissue using the Blood and Tissue DNEasy kit (Qiagen), and 658 bp fragments of the COI gene were amplified and sequenced following the protocol described in . Finally, the resulting dedicated reference DNA database was organized in relevant gene-specific alignments created using the mafft alignment plugin as implemented in Geneious Prime 2020.2.1 (Biomatters Ltd).

### 2.2. PHASE 2 – *IN SILICO* SELECTION AND *IN VITRO* ASSESSMENT OF MOLECULAR ASSAYS

*Assay selection and design (in silico)*

The peer-reviewed scientific literature was queried for existing assays (i.e. PCR/qPCR/dPCR molecular assays targeting relevant marker/taxon combinations) by searching public repositories (i.e. Google Scholar, Web of Science, Pubmed) using key words "PCR", "qPCR", "dPCR" in combination with the common and scientific names of target species, regardless of whether studies used DNA isolated directly from tissue of the target species or environmental samples (e.g. plankton, sediment or water). Subsequently, primer-pairs were aligned to the reference DNA data and manually examined for mismatches with target and off-target species. The same alignments were also used to identify suitable conserved but taxon-specific DNA regions for the design of new primers. The suitability of promising primer sets was tested using Oligoanalyzer (IDT, Coralville, IA, USA), following the criteria detailed in**Table 2** . Where possible, multiple markers per species were designed/chosen to increase species-specific marker development success rate. To further test for assay specificity, primer pairs that satisfied the above criteria were also evaluated using a machine learning tool that accurately predicts qPCR amplification (eDNAssay), as described in The default settings were used and, as a model output, the lowest assignment probability value (AMP value) was reported including the mismatches with that specific target on each primer. Primer pairs that fitted all relevant criteria and had desirable AMP value (<0.5) (where possible) were retained for subsequent *in vitro* testing.

*Assay efficiency, specificity, and sensitivity assessment (in vitro)*

This step was facilitated by the use of a High Throughput qPCR (HT-qPCR) platform (Biomark HD, Standard BioTools Inc, USA), which enabled the time- and cost-efficient *in vitro* assessment of a relatively large number of makers (e.g. compared to conventional qPCR instruments). For this step, template DNA (i.e. extracted from tissue of available specimens) was normalized to 1.5 ng/µl per sample following spectrophotometric quantification with a dsDNA BR (Broad Range) assay kit using a Qubit 3.0 fluorometer (Invitrogen). In addition, to further test specificity of *Mytilus* and *Ensis* species complexes, synthetic double stranded DNA (GBLOCKS, Integrated DNA Technologies) of the relevant target genes was obtained for *Mytilus galloprovincialis, Mytilus edulis, Mytilus trossulus, Ensis ensis, Ensis siliqua, Ensis leei (* synonym *Ensis directus), Ensis magnus (synonym Ensis arcuatus)* and standardized to 10,000 copies/µl. Additionally, to act as a universal positive control and generate standard curves for all assays tested, a six points 10-fold dilution series (ranging from 1,000,000 to 10 copies per µl) was created with an oligonucleotide pool containing synthetic double-stranded DNA fragments of all target genes/species combinations (OLIGOPOOLS, Integrated DNA Technologies)**(** see supplementary material **Table S1)** .

The HT-qPCR runs were performed using IFC controller MX and HX (for priming and loading Integrated Fluidic Controls, IFCs) and a BIOMARK HD real-time PCR system (Standard Bio Tools Inc, USA), following manufacturer's recommendations, including a pre-amplification step (as per protocol QR 100-5876C2) and the actual qPCR runs using the 48.48 IFC (as per protocol QR 100-9791Rev3) or 96.96 IFC configurations (as per protocol QR 100-9792Rev3)(all protocols available at https://www.standardbio.com/support/instrument-support/biomark-ep1-support). Exceptions included the addition of of 1µg/µl BSA (Thermofisher) in the pre-amplification reaction, and subsequent PCR reaction clean up by addition of exonuclease at 4 U/reaction (New England Biolabs) and a thermocycling profile including 37°C for 30 minutes and 80°C for 15 minutes. Cleaned up products were then diluted 10-fold with TE suspension buffer (Qiagen) prior to performing the actual qPCR run on the relevant IFC controller and Biomark HD. Final thermal cycling followed manufacturer's protocols and included a melt curve analysis (i.e. "GE Fast 48x48 PCR+Melt v2.pcl" or "GE Fast 96x96 PCR+Melt v2.pcl"). Each configuration enables the simultaneous real-time qPCR amplification (in parallel) of either 2304 (48 samples tested against 48 assays) or 9216 (96 samples tested against 96 assays) reactions. This allowed the screening of all target species' template DNA in duplicate, the standard dilutions points in six technical replicates for each dilution point per run, including all assays (at least in duplicates) as well as 3 Internal Amplification Controls (IACs) (see supplementary material**Appendix S1** for a more detailed overview of the protocols used). Following each run, qPCR data was inspected using the Real-Time PCR analysis software version 4.5.2 (Standard BioTools Inc, USA). Output data was exported and analysed in R-studio (Posit team, version 2022.12.0+353).

To assess sensitivity of each assay, data generated by the six points 10-fold dilution series was used to

estimate Limit Of Detection (LOD) and Limit Of Quantification (LOQ) for each assay following a discrete threshold as well as a curve-fitting modelling approach, as described in Klymus et al. (2020). The discrete threshold approach for LOD identifies the lowest concentration with 95% positive replicates, while for LOQ the lowest standard concentration that could be quantified with a CV value below 0.35 was selected. These CVs were modeled for exponential decay, linear, and polynomial models to select the lowest residual standard error. When possible, a threshold of CV 0.35 was used on the model to estimate the number of copies for modeled LOQ. Assay efficiency was checked according to expected theoretical values (i.e. optimal efficiency range 90-100% and a regression line slope between -3.6 and -3.3), whereas assay specificity was assessed by checking whether positive amplification (i.e. beyond the LOD) was observed in the target species and/or any off-target species, as well as by melt curve peaks comparison with a difference of >1 degree Celsius selected as the threshold for melt-curve specificity.

After calculating the efficiency, LOD and LOQ, two methods were used to quantify the detected DNA Cq-values to copy numbers. The first method which is the more conventional method, is based on the obtained slope and intercept values from the standard curve. A second method was used to recalibrate intercepts based on the efficiency values of all the assays (see supplementary material **Appendix S2** for a more detailed explanation). Assuming that efficiency was accurately calculated over 5 similar concentrations for all markers in the HT-qPCR, for the second quantification method, the intercepts were plotted against the efficiency of the assays, which resulted in a significant linear model. To improve the fit of the linear equation, outliers were defined based on the standardised residuals of the linear model. More specifically, any observation with a standardised residual equal to or greater than 2 (in absolute value) was deemed to be an outlier and was removed to improve the fit. The linear model was then used to calculate an additional intercept that was used to convert Cq values into copy numbers. The assays specificity was assessed by verifying the positive reactions, grouped per sample/assay, using the melt curve and AMP values. Results were displayed by creating a heatmap showing average amplification strength in copies per/µl, with the corresponding AMP values of the same PCR reaction. Matrices for the heatmaps were created in R-studio with visualization done in Excel (Microsoft, Version 2208).

### 2.3. PHASE 3 –FIELD TESTING AND VALIDATION

*Enzymatic inhibition*

DNA extracted from environmental samples is known to potentially include co-extracted enzymatic inhibitors, which negatively affect the polymerase in PCR and can lead to underestimation of original target DNA copies as well as false negatives . Different strategies have been proposed to mitigate the adverse effects of inhibitors, including the dilution of template DNA used and/or the addition of inhibition-reducing agents such as Bovine Serum Albumin (BSA). Thus, since a dilution-like effect of template DNA is already expected when carrying out qPCR reaction at nano-liter scales (as is the case for droplet digital PCR (ddPCR) as well as the microfluidic-enabled HT-qPCR system used in the present study), the inhibitor-resistant effects of adding BSA to the intended HT-qPCR protocol was tested against two well-known inhibiting substances (i.e. humic acid and EDTA) with the aid of Internal Amplification Control (IAC) assays . For this step, 48.48 IFC configurations were used (using protocol QR 100-9791Rev3), whereby the qPCR amplification was assessed with or without the addition of BSA (at 1 µg/µl concentration). To test for potential effects on different concentrations of target DNA, three distinct IACs were tested at 100000, 10000 and 1000 copies/µl, respectively, whereas to test for the effect of BSA against different concentrations of inhibiting substances, five concentrations of humic acid (2500, 1250, 625, 312.5, 156.25 µM) and five of EDTA (40, 20, 10, 5, 2.5 ng/µl) were tested. All reactions were carried out in triplicate, and concentration ranges of inhibiting substances and evidence of inhibition in the form of delays or absence of amplification in spiked samples (n = 30) vs non-spiked samples (n = 30) were chosen based on previous studies , with HT-qPCR conducted using 48.48 IFCs.

*Field experiment*

To validate the panel of markers using actual samples, zooplankton samples were collected on 21 October

2019 in waters around Howth Head (see **Appendix S3** for further details), where mussel beds are known to occur and expected to be spawning at the time of sampling. Samples were collected at five stations in five replicates per station (n = 25) by means of vertical tows using a 100 µm mesh plankton net with a diameter of 40 cm. Depth and vertical profiles of temperature and salinity were recorded using a YSI Castaway CTD probe (SonTek). Upon collection, samples were equally split into two 50 ml tubes and preserved in Lugol's iodine solution (for microscopical analysis), and 70% ethanol (for DNA analysis), respectively. Microscopic analysis was carried out using a Cobra Micro Zoom MZ1000 stereo microscope, where larvae belonging to the class Bivalvia and Malacostraca were counted, and their relative abundance (larvae/$m^3$) was calculated by dividing the total number of larvae by the approximate volume ($m^3$) of seawater sampled (estimated by multiplying the net opening area by the depth). The 25 ethanol samples were extracted with the Power soil Pro kit (Qiagen), using a customized protocol for plankton samples (see **Appendix S4** for for a detailed protocol). The HT-qPCR run was conducted using 96.96 IFCs configuration and following relevant protocols as indicated above with the inclusion of BSA (see also results section). For quantification the conventional and modeled approach were used to calculate the total copies detected by the HT-qPCR, the copies were averaged per site and compared to the microscopic data. For statistical analysis, ANOVA with Tukey tests were used to determine significant difference between sites for both indirect and direct methods (microscopy and genetic analysis).

## RESULTS

### 3.1. PHASE 1 - DEFINITION OF TARGET TAXA AND ESTABLISHMENT OF REFERENCE DATABASES

Following a review of WoRMS and the Irish Biodiversity Data Centre databases, a total of 151 bivalve and 147 decapod species were determined to occur in Irish waters (**supplementary Table S2** ). Among these species were 24 selected target species as well as 48 bivalve and 18 decapod closely related off-target species (i.e. same Order). The most relevant and appropriate gene commonly targeted for molecular species identification of these taxa was identified as the mitochondrial Cytochrome Oxidase I (COI) with some studies targeting the 16S rRNA gene and the PAPM gene (nuclear) . The COI gene was selected as the main gene of interest for this study as it is more widely studied and better represented via BOLD with more sequences present. While relevant DNA sequence data was available on public databases for all target species, no genetic data were available for 10/48 bivalve and 5/18 decapod species (**supplementary Table S2** ). Of the gathered tissue material, barcoding and relevant DNA sequence data were obtained from 49 samples (encompassing 15 distinct species) (88% successful amplification/sequencing) (**supplementary Table S3** ). Tissue DNA belonging to*Aequipecten opercularis* (queen scallop) and *Pecten maximus*(king scallop) did not yield satisfactory results. As expected,*Mytilus edulis, Mytilus trossulus* and *Mytilus galloprovincialis* were not resolved to the species level using the COI gene. Overall, once previously available and newly generated DNA sequences were included, the dedicated reference database achieved coverage of 80% (for bivalves) and 70% (for decapods) of target species (**supplementary Table S2** ).

### 3.2. PHASE 2 – *IN SILICO* SELECTION AND *IN VITRO* ASSESSMENT OF MOLECULAR ASSAYS

In total 23 different primer pairs were sourced from the literature (14 publications) and 62 additional primer pairs were designed and selected for further *in silico* testing (total n = 85 primer pairs). Following initial assessment of each primer pair, 42 assays were selected for further *in vitro* testing, including 34 assays targeting mitochondrial DNA 16s (n = 3) and COI (n = 31) genes. To gather more information on gene variation and increased taxonomic resolution for *Mytilus spp* and *Magallana gigas* , three cytoplasmic organelle 18s gene assays (n = 3) as well as nuclear genes (n = 5) were added to the selected assays (overview of all assays in**supplementary Table S4** ).

Results from the initial panel of 42 assays showed that average amplicon length was 104 bp and average primer length was 22.5 bp. The average theoretical annealing temperature was 56.2°C, ranging from 52.7 to 61.7°C, with 17/83 primers outside of the 55-60°C range. Delta-G values of the whole panel averaged

-3.9, however, 49/83 primers had one or more reported Delta-G value higher than -6.0 with 26/83 primers having a higher value than -9.0. When blasting the primers, 62/83 were found to match only with the target species. When taking both primers into account, 5 out of the 42 primer pairs matched some non-European off-target species (**supplementary Table S4** ). Two primer pairs matched with European species: ''*M. gigas* COI 2" matched with*C. angulata* (not occurring in Irish coastal waters) and ''*O. edulis* 16S" matched with different *Ostrea* species and was therefore considered as an *Ostrea* specific marker. Using the machine learning approach for predicting cross-amplification (eDNAassay), 32 out of 42 assays were found to have off-target AMP values below 0.5 including the highly specific nuclear *Mytilus spp* markers, hence indicating high specificity potential for 71% of selected assays **(supplementary Table S5)** . The three nuclear assays targeting the *Mytilus* complex were *Mytilus* -specific but had AMP values ranging from 0.5 to 0.675. The results from the COI showed eight assays to have high AMP (>0.5) with off-target species, including five *Ensis* -specific assays (possible cross-amplification with other *Ensis/Pharidae* species; 0.5-0.9 AMP), the *C. glaucum* assay (possible cross amplification with*Parvicardium exiquum* (0.8 AMP)), the *Cancer pagurus* assay (possible cross amplification with *Liocarcinus depurator* (0.6 AMP)) and the *Homarus americanus* assay (possible cross amplification with *Eurynome spinosa* (0.6 AMP)). Despite the potential for cross-amplification, all 42 assays were tested *in vitro* , as described below.

*Assay efficiency and sensitivity*

The *in vitro* assessment of assays was carried out with 2 HT-qPCR runs (96.96 IFC configuration, total 9,216 reactions). The average SD from both runs was 1.8% indicating little variation between the two separate runs (**supplementary Table S6** ). The 42 assays had an average efficiency of 94.5% ($R^2$ ranging from 0.99-1.00), with all assays showing efficiency of 90-102.3%, with the exception of three assays which were found to be below 90%:*Msquinado* COI (87.8%), *Pmaximus* COI1 (84.3%), and*Medulis* PAPM (77.9 %) (**Table 3, supplementary Table S6** ). Most assays showed high detection probabilities with 39 out of 42 assays having a LOD of <10 copies/µL and the remaining three assays at 100 copies/µL, lowering to 10-34 copies/µL when using the modelled approach (**Table 3, supplementary Table S6** ). As for the Limit of Quantification (LOQ) analysis, the discrete results ranged from 10-1000 copies/µL, whereas the modelled approach for the LOQ ranged from 10-387 copies/µL, with 23 out of 42 assays obtaining a 4.4-fold increase in LOQ by the curve-fitting approach over the discrete approach (**Table 3, supplementary Table S6** ). The LOQ could not be determined for 15 assays as it is below 10 copies. There was no evident pattern between lower efficiency, LOQ/LOD and the observed primer variables. Of 42 assays, 33 assays had a LOQ <50 copies/µL, four assays had a LOQ between 50-387 copies/µL, and the five remaining assays had a LOQ <61 copies/µL. When assessing the quantification based on the Standard curve, it appeared that some assays with similar efficiency had different Cq-values, even though the starting quantity was the same for all the targets. After the removal of the outlier assays the efficiency of different assays showed a linear relationship with the intercept. This model *($p < 0.005$, $R2 = 0.7$)* was used to recalibrate the intercept, which was then used as an alternative method to accurately quantify the copy numbers **(supplementary Table S7 and supplementary Appendix S2)** .

*Specificity*

Each species/assay combination (26 *42 = 1092 combinations) yielded at least four qPCR data points (i.e. assay in duplicate tested against the target in duplicate). When assessing assay performance, consistency in amplification across all four qPCR reactions was considered. The overall results are shown in **Figure 2** . In total, separating Bivalvia and Malacostraca, AMP values were generated for 276 reaction combinations. 12 of these reactions showed minor but positive amplification when the AMP value was below 0.3. 3 out of 55 reactions were positive with AMP values of 0.3-0.4 whilst 0 out of 5 reactions with AMP-values of 0.5 amplified. Of the higher AMP-values,*Mytilus edulis* assay reacted positively on the obtained *M. trossulus* specimen from Scotland, however, the obtained oligo based on reference data remained negative resulting in it being considered an artifact. The *M. galloprovincialis* assay amplified *M. edulis* target (100000 copies) <10 copies (AMP=0.7). Following removal of the below 0.2 AMP value reactions (as they were considered artifacts), of the 42 assays tested 38 showed satisfactory levels of specificity to the intended target species,

whereby they either amplified only the target or produced a distinct melt temperature. The 4 remaining assays amplified other targets with similar melt curves, indicating a lack of specificity. Out of these 4 assays, 2 assays*Hamericanus* COI, and *Oedulis* 16S showed substantial amplification in other closely related off targets (signal strength above 10 copies/µL). The other 2 assays (*M. galloprovincialis*(PAPM), *Mytilus* 3 (18S) showed positive but weak amplification (signal strength below 10 copies/µL). Both *Mytilus* and*Ensis* species complexes could be successfully resolved based on inspecting melt curves and cross-referencing amplification (or lack thereof) of multiple assays. Specifically, the *Mytilus* markers on the PAPM gene showed a range of melt curves, including 81°C (*Mytilus edulis* ), 79-80°C (*Mytilus trossulus* ), and 79°C (*Mytilus galloprovincialis* ). Similarly, the *Ensis* species (COI) marker generated distinct melt curves at 75°C (*Ensis siliqua* ) and 77°C (*Ensis ensis* ), hence providing some level of resolution within these species complex. Based on the highest AMP-values tested, detected and the in vitro testing 29 out of 42 assays were found to have no risk of false positives (**Table 3,supplementary Table S6** ).. A total of 6/42 assays were found to have a risk that could be eliminated by another assay present on the panel. Furthermore, a total of 4/42 assays were considered non-specific and therefore the detection of *C. pagarus* and *C. glaucum*cannot be confirmed if the absence of *Liocarcinus depurator* and*Parvicardium exiguum* is uncertain. The remainder of the target species could be successfully detected by the newly developed HT-qPCR panel (**Table 3, supplementary Table S6** ).

### 3.3. PHASE 3 – FIELD TESTING AND VALIDATION

*Testing inhibition*

IPC, within the context of the described triplex assay, was tested against two commonly encountered PCR inhibitors (**Figure 3)** . The reactions without BSA showed almost complete inhibition at 40 ng/µl humic acid, with only the 100.000 copies IPC amplifying with a 17.5 cycle delay. BSA added to the PCR reaction prevented full inhibition with a minor delay being between 0.5-1 cycle for 40 ng/µl, displaying a clear reduction in inhibition. As expected, BSA had less of an effect on EDTA, which caused inhibition for 2500 and 1250 µM in all PCR reactions. The remaining samples tested in this study were all non-inhibited.

*Field validation*

Amplification of internal controls (IACs) showed no observable difference in Cq values between field samples and no template controls (data not shown), indicating no potential inhibition issues and confirming the robustness of the protocol when applied to DNA from actual environmental samples. Screening of collected samples using the direct method (e.g. microscopy), revealed the presence of bivalve larvae in all the sampling sites (averaging 308 larvae per $m^3$of seawater processed per site) and of decapods in all sampling sites except for site 3 (ranging from one to seven individual larvae per site) (**Table 4** ). Across all samples collected, genetic screening using the developed HT-qPCR panel (indirect method) resulted in 13 positive molecular markers detecting the presence of DNA from seven different species, with *Mytilus edulis* being detected in all five sites (**Figure 4** ). Overall, number of larvae and estimated DNA copy numbers (i.e. cumulatively among all markers) showed similar patterns across the five sites for both bivalve and crustaceans, with sites 1, 2 and 5 showing significantly higher larvae/DNA than sites 3 and 4 (p-values varying from 0.0005***-0.05*, Tukey test)**(Table 4).** The copy numbers detected for *Mytilus spp* varied across the three genes used (nuclear PAPM, mitochondrial COI, and cytoplasmic 18S), but were consistent over sites with a nuclear: mitochondrial: cytoplasmic ratio of 1: 16: 78, showing potential to follow expected trends in terms of DNA copies per cell (i.e. 1 nucleus vs many organelles and mitochondria). *Cerastoderma edule* and*aequipecten opercularis* were detected in all sites except for Site 3. The copy numbers of these species varied over the different sites, with higher prevalence of *A. opercularis* in Sites 2 and 5, *C. edule* in Sites 1 and 2. *Mya arenaria* was only consistently present in all replicates of Site 1, but sporadically detected in Sites 3 and 4. *Ensis magnus/ Phaxas pellucidus* was detected in lower presence with most presence in Site 5. The only decapod DNA detected was *Carcinus maenas* in Site 5, which was the site with the highest decapod larvae count observed using the direct method (**Table 4** ). The lack of decapods' DNA detection in Sites 1, 2 and 4 (where decapod larvae were actually observed), implies that these results are either false negatives due to either being below the limit of detection or simply not present in the DNA samples due to very low numbers, and/or the decapod larvae in these samples are not species included in the current molecular HT-qPCR

panel.

## 4 | DISCUSSION

The aim of this study was to avail of a state-of-the-art technology (HT-qPCR) to facilitate the cost- and time-effective development and application of a molecular screening toolkit in a case study to detect a range of bivalve and decapod larvae in zooplankton samples. The overall process was divided into three main phases consisting of the definition of target taxa and establishment of reference databases (PHASE 1);*in silico* selection and *in vitro* assessment of molecular assays (PHASE 2); and field testing and validation (PHASE 3). While PHASE 1 is communal to any general protocol aiming at developing taxon-specific molecular assays, PHASE 2 and 3 are enabled by the HT-qPCR platform. Following initial marker selection, and *in silico* and *in vitro* assessment, the resulting panel of assays included a total of 42 marker/species combinations targeting 24 species of relevance to Irish coastal marine ecosystems and of interest to the local shellfish aquaculture industry. The molecular toolkit (indirect method) performed well when validated on actual zooplankton samples by comparison to a direct method (microscopy), which was used to determine presence and abundance of bivalve and decapod larvae. The following sections address considerations and resulting recommendations of this new approach, which shows great potential in boosting the development of biomonitoring and ecological assessment tools.

*Technical and practical considerations regarding the assay panel development*

This study showed that a HT-qPCR system substantially boosted the number of assays that can be tested *in vitro* (e.g. 96 assays in a single run) resulting in great time-effectiveness and therefore, cost efficiency proposed a 5-level validation scale for determining the readiness of eDNA assays, ranging from "Level 1 – Incomplete" to "Level 5 – Operational". When available, a HT-qPCR system could therefore play a crucial role in transitioning many molecular assays from Level 2 (i.e. testing assays against non-target organisms) to Level 3 and 4 (validating assays in the field), because many assays can be tested in parallel. Level 4 differs from Level 5 as it is missing statistical modelling for detection probability, but both these two last levels are considered valid to provide satisfactory interpretation of presence/absence of target species. Furthermore, the possibility to co-screen a given sample for many species in a single qPCR run would also result in saving precious eDNA template, hence increasing the potential number of targets that can be tested.

Efficiency was satisfactory even though assays were combined into communal thermal cycling conditions, confirming robustness of marker selection steps and reagents used. However, issues with standard for quantification delay in pools of synthetic targets were unexpected. This study experimenting using a second quantification method, which made quantification more accurate but needs further investigation. In future studies comparison between synthetic and pools of genomic DNA of known quantity from target organisms could circumvent these problems of quantification.

Using a machine learning tool (eDNAassay, Kronenberger et al., 2022) during the *in-silico* selection of markers proved to be very useful in providing an initial screening of marker panels and in particular, to assess potential biases due to data gaps in the reference database used. For instance, 17 out of the 42 selected assays showed higher affinity (i.e. higher AMP values) for species not necessarily closely related to the target species, for example, *Ensis* assays with *P. pellucidus,* which combined with the absence of reference data from another closely related Pharidae species, *Pharus legumen* , could lead to potential false positives. This suggests that the specificity of some assays may be lower than expected. Nonetheless, when selected markers were tested *in vitro* , the concordance between expected and observed amplification was good, but in sporadic instances low AMP values still reacted positive. To mitigate, in uncertain situations, multiple markers are preferably selected for various reasons i) to account for the possibility that intraspecific variation prohibits amplification, ii) to increase the target specificity in the case of markers not being fully diagnostic in relation to other species, iii) to quantify the presence of targets more accurately. In addition, multiple markers can also help troubleshoot*in vitro* experiments by identifying possible contamination. This study identified 21 different primer pairs from literature, eventually, only five were used with four being redeveloped as adaptations on existing primers. Most primers from the literature were either developed for a

different purpose or did not fit in the assay requirements. The development of a multi-species panel requires strict rules to make multiple assays work under the same circumstances. Protocol optimization showed that extraction protocol chosen and inclusion of BSA are recommended for genetic screening of zooplankton samples.

Assays developed using the HT-qPCR system are expected to be directly transferable to conventional qPCR instruments ; however various aspects especially related to quantification and sensitivity (e.g. LOD, LOQ) should be re-estimated and confirmed when using a different instrument. Nonetheless, as indicated above, the use of a HT-qPCR system enables the rapid testing of multiple assays in parallel, which is a time-consuming step in assay validation, prior to executing further downstream steps in the laboratory.

*Validation of the molecular toolkit using zooplankton samples*

This study showed good correlation between counts of decapod and bivalve larvae using microscopy (direct method) and strength of genetic signal (indirect method), with the latter providing greater taxonomic resolution power to the species level. Even though previous studies successfully found a similar correlation, it is unclear how the size of larvae and different genetic markers affect signal strength . In this study, the different assays in genes show different results in COI, 18S and nuclear genes, meaning that gene selection is of great importance.

The samples collected for this study were carried out using a 100µm mesh plankton net. Even though such mesh size is not expected to trap smaller fractions of DNA that may be present in the environment such as gametes or other cells released by adults, there was a risk that some target DNA present in a given sample did not belong to larvae (e.g. eDNA adsorbed to larger particles). While this aspect can be challenging to elucidate, the variation of DNA signal in multiple field replicates (n = 5) and the relative proximity of the locations sampled suggested that DNA of target organisms was not ubiquitous and when detected showed differential strength. While discerning fractions of eDNA in plankton samples is beyond the scope of this study, these results emphasize the importance of field replication as well as the usefulness of providing semi quantitative data (e.g. beyond presence-absence. The possibility to incorporate multiple markers/genes targeting the same taxon (species or genus) makes HT-qPCR systems attractive, as this can prove useful in resolving co-occurring closely related species and possibly provide more reliable estimates of DNA copy numbers and hence relative abundance or biomass. For instance, *Mytilus* spp and *Ensis* spp complexes consist of very closely related species in which hybridization might occur. However, single marker approaches, as developed by Inoue et al. (1995) only provide an indication of a genotype but lack the resolution power to clearly distinguish between species. Another benefit of HT-qPCR tool is that some markers can rule out possible false positives. For example, *C. pagurus* and *E. siliqua* are species that cannot be quantified and due to a risk of false positives, including markers for *P. pellucidus* , *P. exiguum* , *L. depurator* could potentially make the panel fully reliable for the species tested.

Finally, this toolkit proved useful in screening plankton samples for the DNA of commercially important species and will be useful in the future for better understanding spawning activities in support of local aquaculture shellfish industry.

*Recommendations for future work and applications*

This study has generated some valuable genetic resources, namely species-specific assays for a range of important species including*Mytilus* spp. The toolkit could be used in future studies if a HT-qPCR platform is accessible, or individual assays could be used using conventional qPCR instruments. The latter would require some further validation. The overall workflow presented here follows general guidelines for the development of biomonitoring HT-qPCR toolkits, encompassing applications such as monitoring of species of conservation concern, surveillance of harmful or unwanted species (harmful algae or invasive species), and supporting fisheries and aquaculture sectors.

Future work includes the continuous improvement of reference DNA databases For instance, in this study, species of scallop (*A. opercularis* and *P. maximus* ) were not successfully barcoded using a general metazoan

protocol , highlighting the need for refining barcoding procedures. Besides taxonomic gaps, public repositories such as GenBank are also known to be hampered by errors , hence flagging the need for the establishment of dedicated and curated databases. On the ecological aspect of this study, current developed methodology needs to be implemented in national routine screening programmes, gathering data over time for the different shellfish species, to generate baseline data and understand distribution and spawning activities. In order to achieve this, a better understanding is needed to translate copy numbers into number of larvae of different species, using different genes Eventually this could lead to fast and cost-effective data collection enabling occupancy modelling or larval dispersal modelling while taking into account life history of species (e.g. duration of planktonic stages)

## ACKNOWLEDGEMENTS

The authors are grateful to Bord Iascaigh Mhara (BIM) for assistance with the project, including sampling aboard the T Burke II survey boat and financial support (Refs: RFT146818, RFT150321). Special thanks go to Terence O'Carroll, Nicholas Chopin, Patricia Daly, and Joanne Gaffney. Financial support was provided by Ireland's European Maritime and Fisheries Fund (EMFF) Operational Program, co-funded by the Irish Government and the EU, as well as from an Irish Research Council postgraduate scholarship (Ref: GOIPG/2021/1712).

## REFERENCES

## DATA ACCESSIBILITY AND BENEFIT SHARING SECTION

### Data Accessibility Statement

Raw data and R scripts are available upon request and safely stored on hard drives, and eventually shared on GITHUB repositories. Raw sequence reads are deposited to NCBI Nucleotide Database (XXXX) including the related metadata can be found in XXX and provides unique sample identifier tags that can be matched to both the deposited genetic data and deposited metadata (for haplotypes, individual sample identifiers and their corresponding haplotype). All samples and reagents are carefully archived in Atlantic Technological Institute in the Marine and Freshwater Research Centre.

### Benefit-Sharing Statement

Benefits Generated: A research collaboration was developed with BIM and ATU and the results of research have been shared with the provider communities and the broader scientific community (see above). This includes the screening tool being adopted and tested on samples collected by local fisherman. In addition, benefits from this research accrue from the shared paper with guidelines that can be followed in future research and experimentation.

### Author Contributions

Dennis van der Pouw Kraan: Conceptualisation, Investigation, Methodology, Lab work, Validation, Writing - original draft, Formal analysis.

Conor T. Graham: Conceptualisation, Writing - review & editing, Supervision.

Fiona Kavanagh: Conceptualisation, Writing - review & editing, Supervision.

Luca Mirimin: Conceptualisation, Investigation, Methodology, Validation, Writing – original draft/review & editing, Supervision, Project administration, Resources,

### Tables and Figures

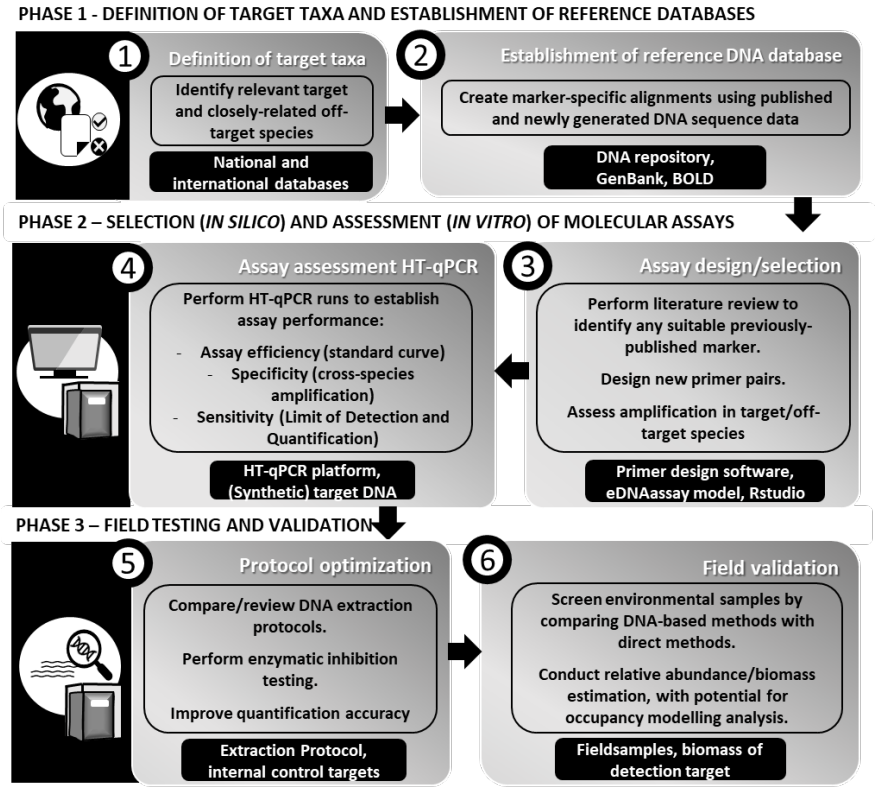



**Figure 1:** Outline and workflow of the methodological approach used in the present study.

**Table 1: Target species (n = 24) identified as main targets in this study. Tissue and/or DNA barcode data availability are also displayed.**

| CLASS | ORDER/INFRAORDER | SPECIES SCIENTIFIC NAME | SPECIES CO |
|---|---|---|---|
| **Bivalvia** | **Mytilida** | *Mytilus galloprovincialis* (Adriatic sea, Italy) | Mediterranean |
| | | *Mytilus edulis* (Killary fjord, Ireland) | Blue mussel |
| | | *Mytilus trossulus* (Scotland) | Bay mussel |
| | | *Mytilus galloprovincialis-edulis hybrid genotype* (Ireland) | Blue mussel |
| | | *Mytilus trossulus-edulis hybrid genotype (*Ireland*)* | Blue mussel |
| | **Ostreida** | *Ostrea edulis* | European Nati |
| | | *Magallana gigas* | Pacific oyster |
| | **Pectinida** | *Pecten maximus* | King scallop |
| | | *Aequipecten opercularis* | Queen scallop |
| | **Cardiida** | *Cerastoderma edule* | Common cockl |
| | | *Cerastoderma glaucum* | the lagoon coc |
| | **Adapedonta** | *Ensis ensis* | (Sword) Razor |
| | | *Ensis siliqua* | Pod razor |
| | | *Ensis leei (Ensis directus)* | Atlantic jackk |
| | | *Ensis magnus (Ensis arcuatus)* | Sword razor cl |
| | **Myida** | *Mya arenaria* | Soft-shell clam |
| | **Venerida** | *Spisula solida* | Surf clam |
| | | *Mulinia lateralis* | dwarf surf clam |
| | | *Rangia cuneata* | wedge clam |

| CLASS | ORDER/INFRAORDER | SPECIES SCIENTIFIC NAME | SPECIES CO |
|---|---|---|---|
| **Malacostraca** | **Brachyura** | *Carcinus maenas* | Littoral crab |
| | | *Maja brachydactyla (Maja squinado)* | Spider crab |
| | | *Cancer pagurus* | Brown crab |
| | | *Necora puber* | Velvet crab |
| | **Caridea** | *Palaemon serratus* | Common praw |
| | **Astacidea** | *Homarus gammarus* | European lobst |
| | | *Nephrops norvegicus* | Dublin Bay pra |

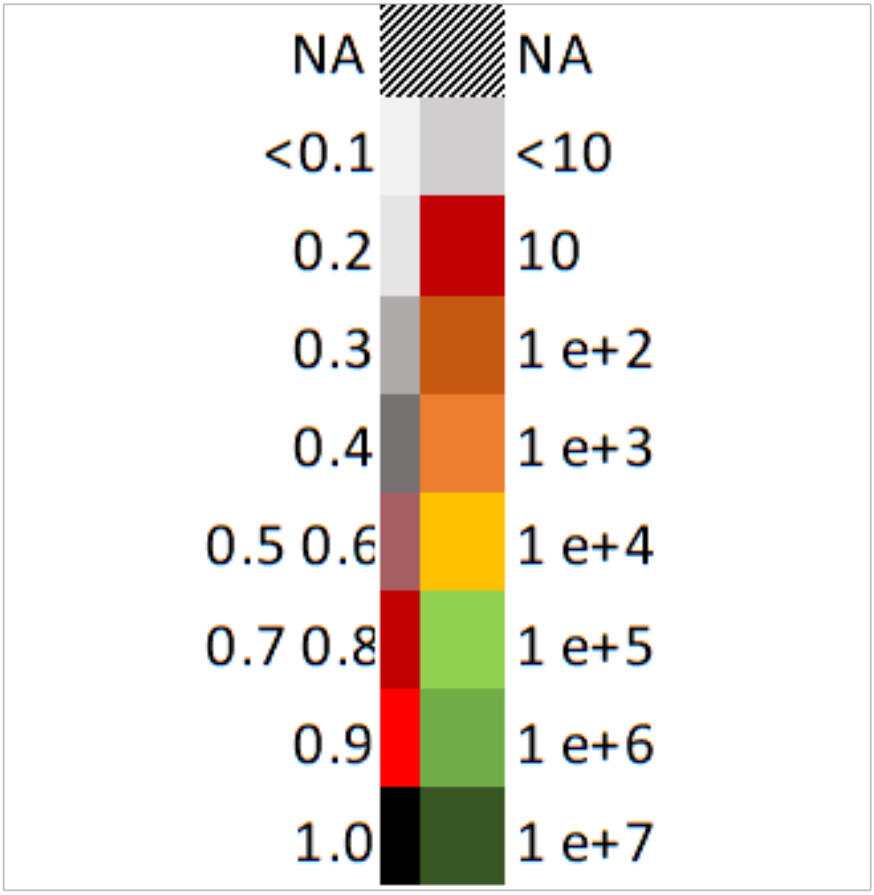
| CRITERIUM # | DESCRIPTION | OPTIMAL RANGE | ACCEPTA |
|---|---|---|---|
| **I** | length fragment | < 100 bp | <300bp |
| | length primers | 17-25 bp | 17-30 bp |
| **II** | Theoretical melt temperature | >55°C - <60°C | >50°C - <65 |
| | range of variation in melt temperature for developed toolkit | < 5 °C | 5 °C - 10 °C |
| **III** | GC% | 40-60 | 30-70 |
| **IV** | ΔG hairpin, self-dimer, hetero dimer | < 6.00 | < 9.00 |
| | bases suitable for binding with own primer sequence | <6 | <10 |
| **V** | BLAST identity | >90% | >99% |
| **VI** | Mismatches both primer site closely related species (Ireland) | >10 | >6 |
| | Mismatches each primer site (within first 5 basepairs on 3') | >3 | >1 |

**Table 2: Criteria used for checking suitability of existing and new primer sets used in PHASE 2.**

**Table 3: Efficiency, sensitivity, and specificity of the newly developed HT-qPCR panel comprising 42 assays.**

| N | Species name | Assay name | Eff (%) | LOD | LOQ | AMP < 0.5 | specific | RISK FALS |
|---|---|---|---|---|---|---|---|---|
| 1 | *A. opercularis* | *Aopercularis*16S | 97.2 | <10 | 11 | Y | Y | NO RISK |
| 2 | *C. edule* | *Cedule*COI1 | 95.1 | <10 | 43 | Y | Y | NO RISK |
| 3 | | *Cedule*COI2 | 96.2 | <10 | 31 | Y | Y | NO RISK |
| 4 | *C. glaucum,* | *Cglaucum*COI | 98.6 | <10 | 10 | N | Y | Parvicardium |
| 5 | *C. maenas* | *Cmaenas_dpk*COI1 | 97.1 | <10 | 15 | Y | Y | NO RISK |
| 6 | | *Cmeanas_crane*COI2 | 92.7 | <10 | 10 | Y | Y | NO RISK |
| 7 | *C. pagurus* | *Cpagurus*COI | 86.8 | <10 | 10 | N | Y | Liocarcinus d |
| 8 | *Ensis spp.* | *Ensisspp COI* | 92.8 | <10 | 32 | N | Y | HIGH RISK I |
| 9 | *E. ensis* | *Eensis*COI1 | 96.2 | <10 | 24 | Y | Y | NO RISK |
| 10 | | *Eensis*COI2 | 96.9 | <10 | 43 | N | Y | Phaxas pellu |
| 11 | *E. siliqua* | *Esiliqua*COI | 97.8 | <10 | 10 | N | N | E. magnus (0 |
| 12 | *E. leei* | *Eleei*COI1 | 96.3 | <10 | 10 | N | Y | Phaxas pellu |
| 13 | | *Eleei*COI2 | 95.0 | <10 | 37 | N | Y | NO RISK |
| 14 | *E. magnus* | *Emagnus*COI | 102.3 | <10 | 15 | N | Y | Phaxas pellu |
| 15 | *H. americanus* | *Hamericanus*COI | 96.8 | <10 | 10 | N | N | Eurynome spi |
| 16 | *H. gammarus* | *Hgammarus*COI0 | 91.3 | <10 | 112 | N | Y | **amplifies H.** |
| 17 | | *Hgammarus*COI1 | 98.4 | <10 | 31 | Y | Y | NO RISK |
| 18 | *M squinado* | *Msquinado* COI | 87.8 | <10 | 22 | Y | N | NO RISK |
| 19 | *Mytilus spp.* | *Mytilusspp*_diasPAPM | 89.6 | <10 | 49 | Y | Y | NO RISK |
| 20 | | *Mytilusspp*_dpkPAPM | 92.1 | <10 | 15 | Y | Y | NO RISK |

| N | Species name | Assay name | Eff (%) | LOD | LOQ | AMP < 0.5 | specific | RISK FALS |
|---|---|---|---|---|---|---|---|---|
| 21 | | *Mytilusspp*18S2 | 94.1 | 34 | 387 | Y | Y | NO RISK |
| 22 | | *Mytilusspp*18S3 | 100.1 | <10 | 12 | Y | N | NO RISK |
| 23 | | *MytilusppCOI1* | 93.3 | <10 | 27 | Y | N | NO RISK |
| 24 | | *MytilusppCOI2* | 95.7 | <10 | 10 | Y | N | NO RISK |
| 25 | *M. edulis* | *MedulisPAPM* | 77.9 | <10 | 61 | N | Y | NO RISK |
| 26 | *M. galloprovincialis* | *MgalloprovincialisPAPM* | 95.5 | <10 | 10 | N | N | M. edulis (0.7 |
| 27 | *M.trossolus* | *MtrossolusPAPM* | 97.5 | <10 | 10 | N | Y | NO RISK |
| 28 | *M.arenaria* | *MtarenariaCOI1* | 96.7 | <10 | 14 | Y | Y | NO RISK |
| 29 | | *MtarenariaCOI3* | 102.1 | <10 | 10 | Y | Y | NO RISK |
| 30 | *M. lateralis* | *MlateralisCOI* | 98.1 | <10 | 10 | Y | Y | NO RISK |
| 31 | *M. gigas* | *Mgigas*_sanchez16S | 93.8 | <10 | 55 | Y | Y | Crassostrea a |
| 32 | | *Mgigas*18S | 91.3 | 14 | 59 | Y | Y | NO RISK |
| 33 | | *Mgigas*COI1 | 97.0 | <10 | 49 | Y | Y | NO RISK |
| 34 | | *Mgigas*COI2 | 93.6 | <10 | 24 | Y | Y | Crassostrea a |
| 35 | *N. norvegicus* | *Nnorvegicus*COI2 | 92.4 | <10 | 10 | Y | Y | NO RISK |
| 36 | *O. edulis* | *Oedulis*16S | 95.3 | <10 | 29 | N | N | M.gigas (0.7) |
| 37 | | *Oedulis*COI1 | 96.2 | <10 | 38 | Y | N | NO RISK |
| 38 | *P. maximus* | *Pmaximus*COI1 | 84.3 | <10 | 10 | Y | Y | NO RISK |
| 39 | | *Pmaximus*COI3 | 90.2 | <10 | 10 | Y | Y | NO RISK |
| 40 | *P. serratus* | *Pserratus*COI1 | 91.0 | <10 | 44 | Y | Y | NO RISK |
| 41 | | *Pserratus*COI2 | 98.4 | <10 | 10 | Y | Y | NO RISK |
| 42 | *R. cuneata* | *Rcuneata*COI | 98.4 | <10 | 94 | Y | Y | NO RISK |




**Hosted file**

`image4.emf` available at https://authorea.com/users/640847/articles/655499-development-and-validation-of-a-dna-based-multi-species-biomonitoring-toolkit-using-a-high-throughput-qpcr-platform-a-case-study-of-irish-shellfish-species

**Figure 2:** Assay performance displayed as strength of amplification expressed as copy numbers (according to scale) (top section of cell with colour and number) and calculated AMP value (bottom section of cell in greyscale). The copy numbers used in the comparison are the modelled copy numbers.

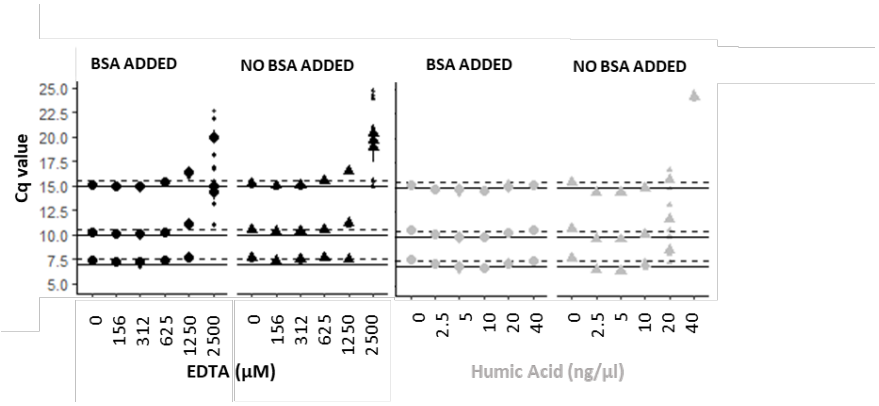


**Figure 3:** Inhibition experiment with 5 inhibitor concentrations for EDTA (**black** ) and Humic acid (**grey** ). The two vertical lines represent inhibition boundaries for each concentration. BSA almost completely prevented inhibition for Humic Acid.

**Table 4** Number and estimated DNA copy number of bivalve and decapod crustacean larvae in the five sampling sites (five field replicate per site) using direct (microscopy) and indirect (genetic) methods. *: presented as the sum of estimated DNA copies/µl over all positive assays. #: values displayed for *Carcinus maenas* COI assay only, as it was the only decapod species that amplified. Significant groups (A-B) and values (>0.05-0.005*, >0.0005***) obtained using a Tukey-test.

| | **Bivalves** | **Bivalves** | **Bivalves** |
|---|---|---|---|
| | **DIRECT METHOD (microscopy) Number of larvae per $m^3$** | **DIRECT METHOD (microscopy) SD (across field reps)** | **DIRECT METHOD (micro Significance group (Tukey-** |
| **Site 1** | 442 | 149 | A* |
| **Site 2** | 592 | 275 | A*** |
| **Site 3** | 37 | 12 | B*** |
| **Site 4** | 99 | 50 | B * |
| **Site 5** | 370 | 162 | A * |

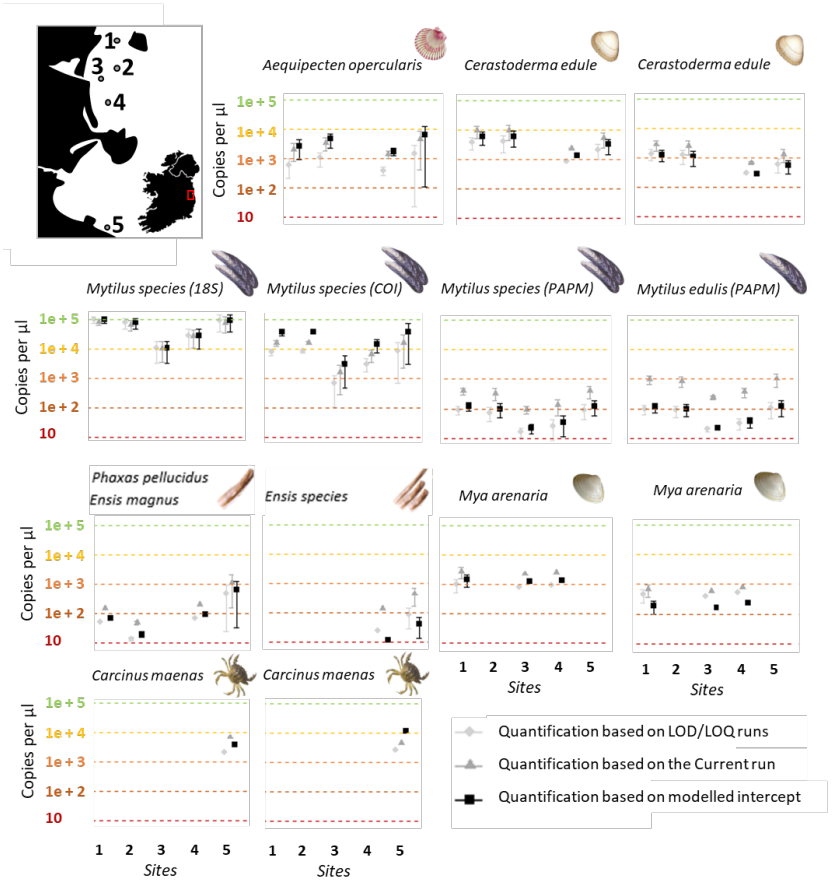



**Figure 4: Field sample analysis with both microscopic analysis and the newly developed HT-**

**qPCR panel.** number of copies detected in each site for each assay using i) quantification based on LOD/LOQ runs for assay optimisation, ii) quantification based on current run st-curve, iii) quantification based on the modelled intercept described.