RESOURCE ARTICLE

A comparison of droplet digital polymerase chain reaction (PCR), quantitative PCR and metabarcoding for species-specific detection in environmental DNA

Susanna A. Wood[1] [iD] | Xavier Pochon[1,2] | Olivier Laroche[1,3] | Ulla von Ammon[1,2] | Janet Adamson[1] | Anastasija Zaiko[1,2]

[1]Coastal and Freshwater Group, Cawthron Institute, Nelson, New Zealand

[2]Institute of Marine Science, University of Auckland, Auckland, New Zealand

[3]Department of Oceanography, School of Ocean and Earth Science and Technology, University of Hawaii at Manoa, Honolulu, HI, USA

Correspondence
Susanna A. Wood, Coastal and Freshwater Group, Cawthron Institute, Nelson, New Zealand.
Email: Susie.wood@cawthron.org.nz

Abstract

Targeted species-specific and community-wide molecular diagnostics tools are being used with increasing frequency to detect invasive or rare species. Few studies have compared the sensitivity and specificity of these approaches. In the present study environmental DNA from 90 filtered seawater and 120 biofouling samples was analyzed with quantitative PCR (qPCR), droplet digital PCR (ddPCR) and metabarcoding targeting the cytochrome c oxidase I (COI) and 18S rRNA genes for the Mediterranean fanworm *Sabella spallanzanii*. The qPCR analyses detected *S. spallanzanii* in 53% of water and 85% of biofouling samples. Using ddPCR *S. spallanzanii* was detected in 61% of water of water and 95% of biofouling samples. There were strong relationships between COI copy numbers determined via qPCR and ddPCR (water $R^2 = 0.81$, $p < .001$, biofouling $R^2 = 0.68$, $p < .001$); however, qPCR copy numbers were on average 125-fold lower than those measured using ddPCR. Using metabarcoding there was higher detection in water samples when targeting the COI (40%) compared to 18S rRNA (5.4%). The difference was less pronounced in biofouling samples (25% COI, 29% 18S rRNA). Occupancy modelling showed that although the occupancy estimate was higher for biofouling samples ($\psi = 1.0$), higher probabilities of detection were derived for water samples. Detection probabilities of ddPCR (1.0) and qPCR (0.93) were nearly double metabarcoding (0.57 to 0.27 marker dependent). Studies that aim to detect specific invasive or rare species in environmental samples should consider using targeted approaches until a detailed understanding of how community and matrix complexity, and primer biases affect metabarcoding data.

KEYWORDS
droplet digital polymerase chain reaction, metabarcoding marine biosecurity, quantitative PCR

1 | INTRODUCTION

The application of molecular based techniques targeting environmental DNA (eDNA) is escalating rapidly and transforming how the Earths biosphere is characterized (Deiner et al., 2017). Most molecular methods used for these purposes can be categorized under the following two main approaches: (a) Targeted methods, which use species-specific primers and techniques such as quantitative polymerase chain reaction (qPCR; e.g., Sigsgaard, Carl, Møller, & Thomsen, 2015; Smith, Wood, Mountfort, & Cary, 2012; Wood,

Zaiko, Richter, Inglis, & Pochon, 2017); and (b) Community-wide characterization methods, which use universal primers that enable entire communities or groups of organisms to be described, e.g., metabarcoding (Thomsen et al., 2012; Wood et al., 2013; Hänfling et al., 2016; Bista et al., 2017). While both targeted and community-wide characterization methods have now been applied widely in a range of habitats (Dowle, Pochon, Banks, Shearer, & Wood, 2016; Evans et al., 2017; Laroche et al., 2016), few studies have directly compared the sensitivity of techniques (Pochon, Bott, Smith, & Wood, 2013). This is particularly important when these methods are applied with the aim of detecting rare or invasive species (Cristescu & Hebert, 2018; Goldberg et al., 2016).

Most targeted methods applied to aquatic systems have used qPCR (e.g., Takahara, Minamoto, & Doi, 2013; Gillum, Jimenez, White, Goldstien, & Gemmell, 2014; Sigsgaard et al., 2015; Wood et al., 2017). Quantitative PCR assays are relatively sensitive and specific, and results can be generated rapidly (<12 hr from DNA extraction to data). However, detection of very low DNA sample concentrations can be challenging to interpret and the reactions, especially when samples are from more complex environments, e.g., biofouling, are prone to inhibition (Opel, Chung, & McCord, 2010). To test for the presence of inhibition most researchers include internal controls, and commonly undertake checks prior to analyzing the actual test samples. Inhibition is usually overcome through the addition of PCR enhancers such as bovine serum albumin (BSA; Kreader, 1996) or through dilution of sample DNA, although dilution is not always advisable when working with low concentrations of DNA. Other nuances of qPCR analysis include the recommendation that samples are analyzed in at least triplicate, and the need to include standard curves that cover at least three orders of magnitude (run in triplicate) in each run when target gene concentrations are required (Bustin et al., 2009). These are generally constructed using DNA, PCR amplicons or plasmids and require additional time and reagents. A more recent advancement to increase the accuracy of standard curves are gBlocks, which are synthetically made sequence-verified, double-stranded DNA fragments.

Droplet digital PCR (ddPCR) offers potential for rapidly expediting this process as no inhibition assays are required, no standard curves are needed, and samples do not need to be run in triplicate as essentially every result is an average of ca. 20,000 individual PCR reactions (Doi, Takahara, et al., 2015; Doi, Uchii, et al., 2015; Te, Chen, & Gin, 2015). Recently, it has been demonstrated that ddPCR can be used to quantify fish eDNA concentrations in mesocosm experiments and field studies (Doi, Uchii, et al., 2015; Nathan, Simmons, Wegleitner, Jerde, & Mahon, 2014). These authors demonstrated that ddPCR is more sensitive than qPCR, especially in samples containing inhibitors.

An increasing number of studies are using metabarcoding to characterize species diversity, or to detect rare or invasive species in environmental samples (Aylagas, Borja, Irigoien, & Rodríguez-Ezpeleta, 2016; Brown, Chain, Zhan, MacIsaac, & Cristescu, 2016; de Vargas et al., 2015; Pochon, Zaiko, Fletcher, Laroche, & Wood, 2017; Zaiko et al., 2016). Unlike targeted approaches, metabarcoding enables many species to be identified simultaneously, and dependent on the approaches used, the taxa may represent multiple trophic levels (Keeley, Wood, & Pochon, 2018). However, many studies have highlighted challenges related to the amplification of multi-template sequences (Zinger et al., 2019) including: primers that are not truly universal, primer biases (Clarke, Soubrier, Weyrich, & Cooper, 2014; Deagle, Jarman, Coissac, Pompanon, & Taberlet, 2014), "tag jumps", which occur when the tag assigned to one particular sample is recombined to sequences belonging to another sample (Taberlet, Bonin, Zinger, & Coissac, 2018), different results based on choice of bioinformatic pipelines (von Ammon et al., 2018), and incomplete or inaccurate reference databases resulting in misassignment of taxa (Dowle et al., 2016; Nilsson et al., 2006; Wangensteen, Palacín, Guardiola, & Turon, 2018).

This study focuses on detection of a marine nuisance species, the Mediterranean fanworm *Sabella spallanzanii* (Gmelin, 1791) (Polychaeta: Sabellidae). It is a large, tube-dwelling polychaete worm native to the Mediterranean Sea and Atlantic coast of Europe (Patti & Gambi, 2001). It was first detected in New Zealand in 2008 (Read, Inglis, Stratford, & Ahyong, 2011), and has since been identified at multiple locations, including several where it is now well established (for specific distribution data refer to https://www.marinebiosecurity.org.nz). Once established, it can form dense populations covering a variety of marine habitats (e.g., Holloway & Keough, 2002).In order to study the efficacy of different detection methodologies for *S. spallanzanii*, water and biofouling samples were collected from Auckland harbour, where *S. spallanzanii* populations can reach extremely high densities (e.g., 100–600 individuals per m^2; Wood et al., 2017). The aims of the study were to evaluate: (a) correlations between qPCR and ddPCR results, (b) the sensitivity of targeted and metabarcoding methods, and (c) determine detection probabilities between methods and sample types.

2 | MATERIALS AND METHODS

2.1 | Sample collection

Thirty water samples were collected in triplicate from ports, marinas and surrounding coastline in Auckland on 19 and 20 April 2017 as described in Wood et al. (2018) (Figure S1). Briefly, the water depth at each site was measured, and 15 L of seawater from three depths: 1 m above the seafloor, 1 m below the water surface, and mid-way between these two depths (total volume = 45 L) was pumped (5.0 GPM Washdown Pump; Seaflow) through the same prefilter (20 μm netting). The water from the three depths was combined for the analysis. Seawater (ca. 50 ml) from the site was added to the filtrate captured on the mesh and was filtered a second time (sterile Whatman GF/C, c. 1.6 μm pore size, 47 mm). The filter was halved using sterile scissors and half filters were placed into two sterile Eppendorf tubes containing LifeGuard Soil Preservation Solution (1 ml; Qiagen) and within 36 hr stored at −80°C for later DNA extraction. Negative controls (1 L tap water) were filtered in triplicate on GF/C filters as described above at every tenth site. Prior to sampling at each site, all sampling equipment was thoroughly washed

—WILEY |

using 5% bleach (sodium hypochlorite) solution for at least 5 min and rinsed in seawater from the sampling site.

Settlement plate arrays were deployed in the Westhaven Marina, Auckland, New Zealand during winter (June to October 2015, n = 5) and summer (November 2015 to February 2016, n = 5). Each settlement array comprised 12 polyvinyl chloride settlement plates with three types of antifouling treatment applied, as described in von Ammon et al. (2018) and Tait, Inglis, and Seaward (2018). At the end of the deployment period, all biofouling material was removed from the plate surface using sterile stainless-steel surgical blades (Swann-Morton) in February (summer) or sterilized sponges (Whirl-pa, Speci-sponge) in October (winter). These were kept on ice and then stored at –70°C in sterile tubes (summer samples) or individual sterile plastic bags (winter samples) until further processing.

2.2 | Laboratory quality control

Each step of the molecular analysis (i.e., DNA extraction, qPCR and ddPCR setup, template addition and qPCR/ddPCR analyses) was conducted in a separate sterile laboratory dedicated to that step with sequential workflow to ensure no cross-contamination. Each room was equipped with ultraviolet sterilization which was switched on for a minimum of 15 min before and after each use. The PCR setup and template addition was undertaken in laminar flow cabinets with HEPA filtration. Aerosol barrier tips (Axygen BioScience) were used throughout.

2.3 | DNA extraction

The tubes containing the water samples were centrifuged (3,000 g, 2 min) and the supernatant removed. The filter was transferred to the BashingBead Lysis Tube which already contained the beads (2.0 mm; Zymo Research) using sterile forceps and any remaining pellet was mixed with lysis buffer from the ZR-Duet DNA/RNA MiniPrep Kit (Zymo Research) and then transferred to the same tube. The tubes were then placed on a bead beater (1600MiniG Spex SamplePrep) for 2 min at 1500 RPM. DNA was extracted from filters using the ZR-Duet DNA/RNA MiniPrep Kit (Zymo Research), according to the manufacturer's protocol. The DNA was eluted in 60 µl of elution buffer provided with the kit.

The total weight of the summer biofouling samples (wet weight) was obtained by first weighing a 50 ml Falcon tube and then adding sample material to this. Each sample was homogenized in a tube (1600MiniG Spex SamplePrep, 2 min, 1,500 rpm). A subsample (ca. 0.2 g) was then weighted directly into the first tube of a DNasy PowerSoil DNA Isolation Kit (Qiagen). The sponges with biofouling samples (winter samples) were macerated in a Stomacher 400 Circulator lab paddle blender (Colworth 400; AJ Seward) for 2 min at maximum speed, then squeezed applying consistent pressure to remove excess liquid. The resulting suspensions were pelleted by centrifugation (4,000 g, 15 min), the supernatant discarded and the pellet transferred to the first tube of a PowerSoil DNA Isolation Kit. (Qiagen) DNA was extracted from the resulting pellet using the PowerSoil DNA Isolation Kit (Qiagen) following the manufacturer's protocol. The DNA was eluted in 100 µl of elution buffer provided with the kit. A negative DNA extract control was included every twenty-third sample.

The quality and purity of isolated DNA were checked using a BioPhotometer (Eppendorf) following the manufacturer's protocol.

2.4 | Quantitative polymerase chain reaction

All samples were screened in duplicate for inhibition using an internal inhibition control assay. Each 10 µl reaction contained 6.25 µl KAPA Probe Fast qPCR Kit Master Mix (2×), 0.5 µl of primers targeting the Internal transcribed spacer region 2 of the rRNA gene operon of *Oncorhynchus keta* salmon sperm (10 µM, Sketa F2 (5´-GGTTTCCGCAGCTGGG-3´) and Sketa R3 (5´-CCGAGCCGTCCTGGTCTA-3'), IDT, USA, Haugland, Siefring, Wymer, Brenner, & Dufour, 2005), 0.2 µl TaqMan probe (10 µM; 5´-AGTCGCAGGCGGCCACCGT-3') labelled at the 5' end with the fluorescent reporter dye FAM-6-carboxyfluorecein and at the 3' end with a nonfluorescent quencher with the Black Hole Quencher-1 (IDT, USA), 0.8 µl DNA/RNA-free water (Life Technologies), 0.75 µl BSA (0.2 mg/ml; Sigma), 1 µl extracted salmon sperm DNA (15 ng; Sigma) and 1 µl of template DNA. The cycling profile was: 95°C for 3 min, followed by 40 cycles at 95°C for 3 s and 60°C for 20 s. All samples showed some inhibition (a shift of 2–3 cycle thresholds from the positive control) and were diluted 1:10 with DNA/RNA-free water (Life Technologies) and re-analyzed for inhibition as described above.

Each sample was then analyzed in triplicate for *S. spallanzanii* using the qPCR assay described in Wood et al. (2017), which targets the mitochondrial cytochrome c oxidase I (COI) gene. Each specific reaction consisted of 10 µl containing: 6.25 µl KAPA Probe Fast QPCR Kit Master Mix (2×), 0.4 µl of the primers Sab3-QPCR-*F* (5´-GCTCTTATTAGGCTCTGTGTTTG-3´) and Sab3-QPCR-R (5´-CCTCTATGTCCAACTCCTCTTG-3´) for *S. spallanzanii* (10 µM, IDT; Wood et al., 2017), 0.3 µl TaqMan probe synthesized with a FAM reporter dye at the 5´end and a Black Hole Quencher-2 at the 3´end (Sab3-QPCR-Probe (5´-AAATAGTTCATCCCGTCCCTGCCC- 3´), 10 µM, IDT; Wood et al., 2017), 0.75 µl BSA (0.2 mg/ml; Sigma), 0.8 µl DNA/RNA-free water (Life Technologies) and 1 µl of template DNA. The cycling profile was: 95°C for 3 min, followed by 40 cycles at 95°C for 3 s and 60°C for 20 s. Five-point standard curves ranging from 7.56×10^2 to 10^6 gene copies per µl and no template controls were analyzed in triplicate on each qPCR run. The standard curve was constructed using a purified (AxyPrep PCR Clean-up Kit; Axygen Biosciences) PCR product generated using the *S. spallanzanii* primers described above. The DNA used to generate this PCR product was sourced from *S. spallanzanii* tissue collected from Lyttelton (New Zealand) in June 2009 (Wood et al., 2017). The number of copies in the PCR product used for the standard curves were determined using: $(A \times 6.022 \times 10^{23})/ (B \times 1 \times 10^9 \times 650)$, with A being the concentration of the PCR product, 6.022×10^{23} (Avogadro's number), B being the length of the PCR product, 1×10^9 used to convert

to ng, and 650 the average molecular weight per base pair (bp). The standard curves were linear ($R^2 > 0.98$) and qPCR efficiency ranged between 0.80 to 0.96.

2.5 | Droplet digital polymerase chain reaction

Droplet digital PCR was undertaken using a BioRad QX200 system. Each ddPCR reaction included 450 nM of each primer and probe (as described above for qPCR), 1 × BioRad ddPCR Supermix for probes, 1 µl DNA, and sterile water for a total reaction volume of 22 µl. The BioRad QX200 droplet generator partitioned each reaction mixture into nanodroplets by combining 20 µl of the reaction mixture with 70 µl of BioRad droplet oil. After processing, this resulted in a total nanodroplet volume of 40 µl, which was transferred to a PCR plate for amplification using the following cycling protocol: hold at 95°C for 10 min, 40 cycles of 94°C for 30 s, 60°C for 1 min, and a final enzyme deactivation step at 98°C for 10 min. The plate was then analyzed on the QX200 instrument. For each ddPCR plate run, at least one negative control (containing all reagents and RNA/DNA-free water) and one positive control (genomic DNA extracted from a sample known to contain high copy numbers of the *S. spallanzanii* COI gene:; sample 49 from Westhaven marina as described in Wood et al., 2018) were included. The results were converted to copies per sample using the following formula: number of copies per µl × 22 µl (the initial volume of the PCR reaction) × the volume used to elute the DNA during extraction.

2.6 | Metabarcoding

For metabarcoding, a segment (approximately 400 bp) of the V4 region of the nuclear small subunit ribosomal (18S rRNA) gene and an approximately 300 bp fragment of COI were amplified by PCR. For the 18S rRNA gene, the eukaryotic-specific primers were Uni18SF: 5'-AGG GCA AKY CTG GTG CCA GC-3' and Uni18SR: 5'-GRC GGT ATC TRA TCG YCT/T3' (Zhan et al., 2013), modified to include Illumina overhang adaptors following von Ammon et al. (2017). Thermocycling PCR conditions were: 94°C for 3 min followed by 35 cycles of 94°C for 30 s, 52°C for 30 s, 72°C for 1 min, with a final extension step at 72°C for 7 min. For the COI gene, the eukaryotic-specific primers mlCOIintF: 5'-GGW ACW GGW TGA ACW GTW TAY CCY CC-3' and jgHCO2198: 5'-TAI ACY TCI GGR TGI CCR AAR AAY CA-3' were used following Leray et al. (2013). PCR amplification was undertaken in a total volume of 50 µl using 25 µl of MyTa Red Mix (Bioline), 2 µl of each primer, 16 µl of DNA-free water, 3 µl of BSA (0.2 mg/ml; Sigma) and 2 µl of template DNA. Thermocycling conditions were: 95°C for 3 min, followed by 40 cycles of 95°C for 30 s, 50°C for 30 s, 72°C for 90 s, and a final extension of 72°C for 10 min. Two samples of 20 µl of RNA/DNA-free water were used as negative sequencing controls following the same protocol as described above.

Purification and quantification of amplicons were performed following the Agencourt AMPure XP protocol (Beckman Coulter), using magnetic beads and a QUBIT 2.0 Fluorometer (Invitrogen).

Purified amplicons were diluted to 3 ng/µl and libraries sent to New Zealand Genomics Limited at the University of Auckland for indexing (Nextera XT Index Kit v2) and sequencing. Paired-end sequences (2 × 250) were generated on a MiSeq instrument using the TruSeq SBS kit (Illumina).

2.7 | Bioinformatics

Raw sequence reads in fastq format were demultiplexed and primers removed using FASTQ-MULTX (Version 1.3.1) from EA-UTILS package (Aronesty, 2011) and reads paired with SOLEXAQA++ (Cox, Peterson, & Biggs, 2010). Reads were truncated from the first base where the Phred score dropped below three, and merged using the software package VSEARCH (Rognes, Flouri, Nichols, Quince, & Mahé, 2016), allowing a total of five mismatches in the overlapping region. Merged sequences were quality filtered on the expected error value using VSEARCH default value of one. Sequence chimeras were detected and removed using the Quantitative Insights Into Microbial Ecology (QIIME) package (Caporaso et al., 2010) and the reference-based method in Usearch (Edgar, 2010), mapping unique 18S rRNA sequences against the Protist Ribosomal Reference database (PR2 database; Guillou et al., 2013), and COI sequences against a custom reference database combining unique sequences from the Barcode Of Life Database (BOLD; Ratnasingham & Hebert, 2007) and the MIDORI database (Machida, Leray, Ho, & Knowlton, 2017). Sequence reads were clustered into operational taxonomic units (OTUs) using the Swarm clustering method (Mahé, Rognes, Quince, de Vargas, & Dunthorn, 2014), with a local clustering threshold (d-value) of three. Taxonomy was assigned using the RDP classifier (Wang, Garrity, Tiedje, & Cole, 2007) implemented in QIIME, trained on the PR2 database for 18S rRNA and the aforementioned custom database for COI.

Raw sequences were submitted to the National Center for Biotechnology Information Sequence Read Archive (SRA) under the accession number PRJNA477646 (water), PRJNA478269 (COI biofouling), and PRJEB25036 (18S rRNA biofouling).

2.8 | Statistics

Data from summer and winter biofouling samples were combined for all analyses. The detection of *S. spallanzanii* among the different methodologies was visualized with a Venn diagram using VENNY 2.1 (Oliveros, 2015). The relationship between qPCR and ddPCR detection of *S. spallanzanii* was investigated with least squares regression with the lm function implemented in R (Version 3.4.3; R Development Core Team, 2014). Upon inspection of the diagnostic plots, two outlier samples (BioF_76, BioF_80 – both summer biofouling samples) were removed to meet the assumptions of the regression. A one-way analysis of variance (ANOVA) was performed to test for significant difference between water and biofouling samples for both ddPCR and qPCR results.

Overall, taxonomic composition of the metabarcoding 18S rRNA and COI data for both sample types (water and biofouling) was visualized at the phylum level with pie charts, using the ten most

TABLE 1 Averaged estimates from the top-ranked model, showing the following parameters: large-scale probability of occupancy (ψ), small-scale probability of occupancy (θ), detection probability (p). Where applicable, standard errors (SE) are given

	qPCR	ddPCR	COI	18S rRNA
Water				
ψ (SE)			0.74 (0.1)	
θ (SE)			0.91 (0.02)	
p (SE)	0.93 (0.02)	1.0	0.57 (0.05)	0.27 (0.05)
Biofouling				
ψ (SE)			1.0	
θ (SE)			0.91 (0.02)	
p (SE)	0.87 (0.03)	1.0	0.40 (0.04)	0.16 (0.03)

Note: Detection probabilities may vary among method (qPCR, ddPCR, COI and 18S rRNA) and both probability of occupancy and detection may vary between sample types (water and biofouling).
Abbreviations: COI, cytochrome c oxidase I; ddPCR, droplet digital PCR; qPCR, quantitative PCR.

abundant phyla present in each data set. Using samples rarefied at 5,000 reads, significant differences in overall OTU richness between sample source and target gene was tested with a Welch's two-sample t test with the R software.

2.9 | Occupancy modelling

To compare detection probabilities of applied molecular methods (qPCR, ddPCR, COI metabarcoding and 18S rRNA metabarcoding), an occupancy estimation modelling approach was employed (MacKenzie et al., 2002). Occupancy estimation and modelling based on detection/nondetection data provides an effective way of assessing species' distribution across time and space in cases where the species is not always detected with certainty (Nichols et al., 2008). This approach is a likelihood-based method for estimating the proportion of sites occupied by a species (occupancy) when detection probabilities are less than one, i.e., occupied sites may be classified as unoccupied based on survey data (Guillera-Arroita, Ridout, & Morgan, 2010). Recently, site occupancy detection models, which provide an estimate of the large-scale occupancy whilst accounting for imperfect detection, have been advocated as a way to overcome the potential issues of interpreting eDNA-based data during field surveys (imperfect detection, patchy distribution of eDNA in environment, imperfect quantification, Lugg, Griffiths, van Rooyen, Weeks, & Tingley, 2017).

The occupancy modelling in this study was based on detection histories of *S. spallanzanii* derived from water and biofouling samples. Although this approach is primarily aimed at estimating the proportion of area occupied by a species when detection probabilities are < 1 (MacKenzie et al., 2002), it also allows quantifying and comparing detection by different methods using repeated site surveys data (Pregler, Vokoun, Jensen, & Hagstrom, 2015). The "multi-method, single season" model variant, implemented in PRESENCE v12.7

(Hines, 2006), was used to estimate detection probabilities for each method.

In the model, water sampling sites ($n = 30$) were treated as sampling units and sample replicates as the repeated surveys. For the biofouling samples, combination of array-antifouling treatment was considered as a sampling unit, and plate replicates as the repeated surveys. Since we were focusing at the method efficiency in relation to two sample types, the effect of antifouling was not specifically addressed in this study, but was considered elsewhere (von Ammon et al., 2018; Tait et al., 2018). When analyzing the metabarcoding data (both COI and 18S rRNA), only "true detections" were considered, i.e. *S. spallanzanii* detections confirmed by either qPCR or ddPCR detections.

The following parameters were estimated:

- $\psi_{\text{naïve}}$ = naïve estimate of occupancy probability, calculated as proportion of sample units where the species was detected over all units surveyed;
- ψ = model-estimated large-scale occupancy, probability of sample unit being occupied by a species;
- θ_t = small-scale occupancy estimate, the probability that the species is present in the direct vicinity of the gear (in this study this could be interpreted as *S. spallanzanii* DNA present in the sample) at occasion t, given that sample unit is occupied;
- p_t^m = probability of *S. spallanzanii* detection at occasion t by method m, given that sample unit is occupied, and species (in this case – DNA of *S. spallanzanii*) is present at immediate sampling site.

A small set of a priori models (Table 1) were defined with varying assumptions to describe the detection data and address the defined research questions. Type of sample (water or biofouling) was used as a covariate, to test its effect on species occupancy and detection probabilities. All fitted models were ranked according to Akaike Information Criterion (AIC) values calculated in PRESENCE software.

2.10 | Receiver operator characteristics analysis

Receiver operator characteristics (ROC) analysis was applied to the metabarcoding outputs (total number of sequence reads and number of *S. spallanzanii* reads) to assess their predictive capacity for targeted detection and to establish threshold values optimized for best prediction. Receiver operator characteristics analyses generate a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied. A ROC graph plots the probability of the true positives (sensitivity) against the probability of the true negatives (specificity). The area under the ROC curve (AUC) can be used as a measure of the response to the predictor. A perfect predictor should have an AUC of 100%, whereas an AUC of 50% is a noninformative predictor (Murtaugh, 1996). In environmental studies, AUC values > 80% are generally considered to indicate an excellent response and > 70% an acceptable response (Hale & Heltche, 2008). Based on the combination of
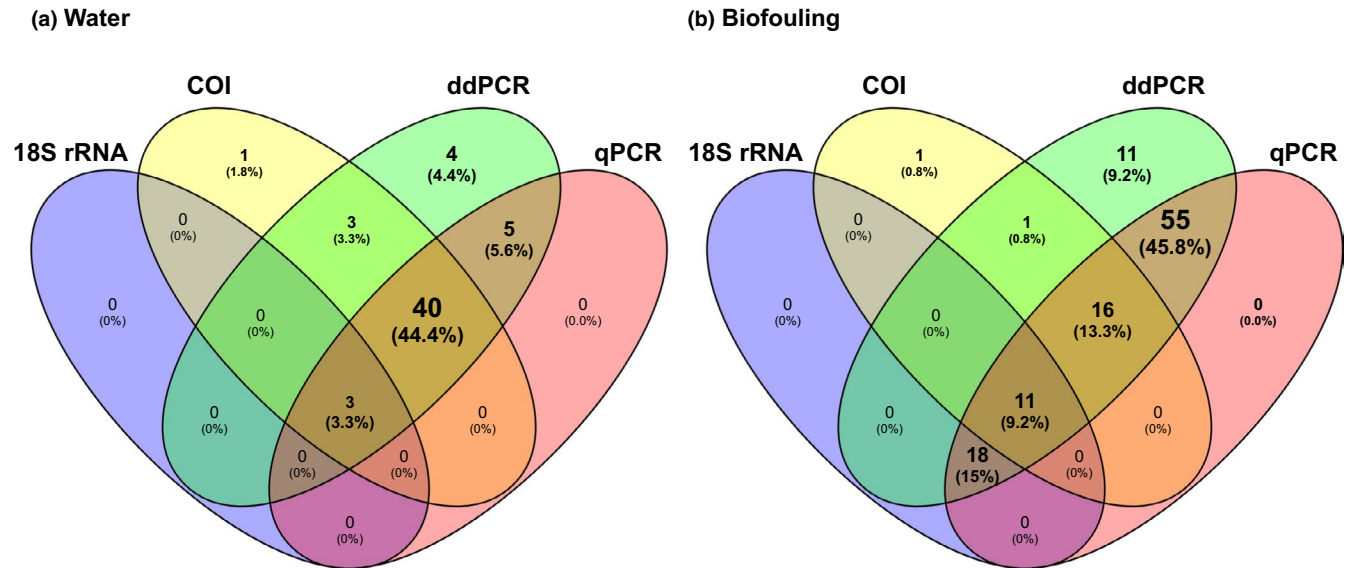
(a) Water

(b) Biofouling

FIGURE 1 Venn diagrams showing the number and proportion of *Sabella spallanzanii* detections for (a) water and (b) biofouling samples using the different methodologies. COI, cytochrome c oxidase I. *n* = 90 for water sample, *n* = 120 for biofouling samples [Colour figure can be viewed at wileyonlinelibrary.com]

sensitivity and specificity values, the probability of obtaining a correct response can be assessed and used to set the optimal threshold for the predictor.

The following hypotheses were tested by considering different "predictor-response" combinations:

1. Sequencing depth affects the probability of correct *S. spallanzanii* detection using metabarcoding. The analysis was run on a subset of samples where *S. spallanzanii* was detected by either qPCR or ddPCR – considered as the "true detection" data. The total number of high-quality sequences from a sample (sequencing depth) was used as a predictor and *S. spallanzanii* detections in metabarcoding data as a binary classifier, for COI and 18S rRNA data, and water and biofouling samples, respectively.

2. Number of *S. spallanzanii* sequences in metabarcoding data affects the probability of correct *S. spallanzanii* detection. The analysis was run on the entire data set, with the number of sequences assigned to *S. spallanzanii* in the metabarcoding data used as a predictor, and *S. spallanzanii* detections by either real time PCR or ddPCR as a binary classifier, for COI and 18S rRNA data, and water and biofouling samples, respectively.

Receiver operator characteristics curves were produced, and AUC values were calculated using the "pROC" package in R (Robin et al., 2011), with a test for a partial AUC undertaken for each "predictor-response" combination (no partial AUCs were detected in any case). The AUCs derived for water and biofouling samples were compared using nonparametric DeLong's test (DeLong, DeLong, & Clarke-Pearson, 1988) implemented in "pROC" package. The optimal threshold for each predictor was estimated as the point closest to the top-left part of the plot (i.e., perfect sensitivity and specificity combination).

3 | RESULTS

There was no amplification from any of the negative controls (sampling or extraction) in either the qPCR or ddPCR. The qPCR analyses detected *S. spallanzanii* in 53% of water and 84.8% of biofouling samples (Figure 1). There were an additional 8% detection using ddPCR in water samples and 10% in biofouling samples. There were no samples where qPCR detected *S. spallanzanii* and ddPCR did not. The COI gene showed a much higher detection rate in the water samples (52%) than the 18S rRNA gene (3%), while detection between COI and 18S rRNA was similar with the biofouling samples (27% and 24%, respectively). Only two *S. spallanzanii* detections with metabarcoding (sample FW_24 [25 COI reads] and sample BioF_74 [2 COI reads]), were not detected with the ddPCR or qPCR methods.

There were strong significant relationships between *S. spallanzanii* copies determined via qPCR and ddPCR among both sample types (water R^2 = 0.81, *p*-value < .001, biofouling R^2 = 0.68, *p*-value < .001; Figure 2). The copy numbers determined by qPCR were on average 125-fold (range 0.01 to 13,000-fold) lower than those measured using ddPCR (Figure 2).

3.1 | Metabarcoding

Two *Sabella* species were detected (*S. spallanzanii* and *Sabella pavonina*) and some *Sabella* sequences were not assigned beyond the genus level. Phylogenetic trees containing all *Sabella* sequences were constructed, and these indicated that the target gene could not distinguish these at a species level. To the best of our knowledge, *S. pavonina* has never been observed in New Zealand. Therefore, all sequences assigned to *Sabella* where combined and are hereafter referred to as *S. spallanzanii*.
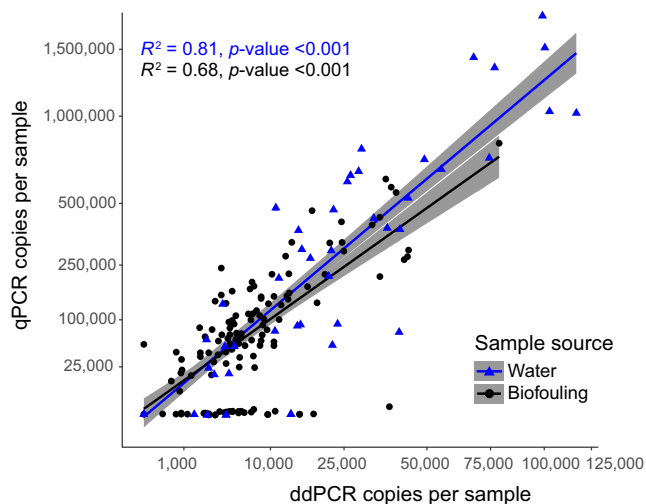
FIGURE 2 Linear regression between quantitative PCR (qPCR) and digital droplet PCR (ddPCR) detection methods for the water and biofouling samples [Colour figure can be viewed at wileyonlinelibrary.com]

From the 90 water samples and 120 biofouling samples sequenced for both 18S rRNA and COI, three biofouling samples failed to amplify for the COI gene (Table S1). The total number of paired-end, quality filtered and nonchimeric COI sequences obtained from water and biofouling samples, respectively, were 2,850,178 reads (mean of 31,669) and 3,532,950 reads (mean of 30,196; Table S1). This corresponded to 42,259 and 45,619 OTUs, respectively. The values for the 18S rRNA gene region were 439,791 reads (mean of 4,941) from water samples and 3,685,649 reads (mean of 30,714) from biofouling samples (Table S1), resulting in 4,737 and 18,810 OTUs, respectively. After quality and chimeras filtering, the two negative DNA extraction controls yielded a total of 33 and nine reads, none of which belonged to *S. spallanzanii*. All other negative controls contained no reads.

In the filtered water samples, the majority of assigned taxa were Arthropoda (18S rRNA = 51% and COI = 19%), Mollusca (18S rRNA = 11% and COI = 3%) and Annelida (18S rRNA = 7% and COI = 2%). For both marker genes, a large proportion of sequences remained either unidentified Metazoan (25%; 18S rRNA) or Eukaryotes (74%; COI; Figure S2a). In the biofouling samples, the taxonomic composition of 18S rRNA data was mostly composed of Urochordata (27%), Bryozoa (25%), Ciliophora (11%) and Arthropoda (11%). This differed markedly from the COI data where, with the exception of Arthropoda (12%), these taxa were found in low abundance. A large proportion (78%) of COI sequences were unidentified Eukaryotes, whereas only 4% of 18S rRNA sequences were unidentified Metazoan (Figure S2b).

When the data were rarefied to an equivalent sampling depth (reads = 5,000), OTU richness in the COI data was significantly lower (*t* test, *p*-value < 0.001) in water samples (mean of 220 OTUs per sample) compared to biofouling samples (mean of 314 OTUs per sample; Figure S3). For the 18S rRNA gene, a marginally significant difference (*t* test, *p*-value = .04) was observed with a mean number of 142 and 125 of OTUs per sample for water and biofouling samples, respectively (Figure S3). Richness of the 18S rRNA OTUs was

also significantly lower than COI for both water (*p*-value < 0.001) and biofouling samples (*p*-value < .001; Figure S3).

3.2 | Occupancy and detection probabilities

Naïve estimates of large-scale occupancy ($\psi_{\text{naïve}}$) were higher for targeted methods (0.85 and 0.87 for qPCR and ddPCR correspondingly) comparing to metabarcoding results (0.65 and 0.33 for COI and 18S rRNA correspondingly). Within the candidate set of models, those assuming constant small-scale occupancy (θ) and method-dependent probability (p) of detection were consistently ranked higher based on Akaike weight values (Table S1). This suggests very strong evidence that detection probability is method dependent.

The most supported model structure accounted for sample type-dependent large-scale occupancy (ψ) and differences in detection probabilities between methods and sample types. All parameter estimates yielded good precision in the top-ranked models, as evidenced by the standard error values (never exceeded 30% of the estimate value, Table 1). Although occupancy estimate was higher for biofouling samples (ψ = 1.0), higher probabilities of detection were derived for water samples. Overall, targeted methods yielded higher detection probability estimates compared to metabarcoding, with ddPCR achieving detection probability estimator p = 1 in both sample types. The lowest detection probability estimates were reported for 18S rRNA-metabarcoding.

Input files for the occupancy modelling are given in Table S2.

3.3 | Receiver operator characteristics

The results of the ROC analysis showed that the effectiveness of "sequencing depth" as a predictor of *S. spallanzanii* detections was higher for water samples for both the COI and 18S rRNA data. The best sensitivity-specificity combination was achieved with lower numbers of reads (1,568 and 4,906, respectively; Figure 3). However, the true difference of derived AUC values was not statistically significant (DeLong's test, *p* = .15 and .09 for COI and 18S rRNA, respectively).

When using number of *S. spallanzanii* sequences as a predictor, an excellent response was derived only for COI data from water samples (Figure 4) and this was significantly different from the non-informative response from the biofouling samples (DeLong's test, *p* < .001). This indicates that true detections of *S. spallanzanii* can be expected if the number of putative target sequences in the metabarcoding is > 0. However, the probability of deriving "true positives" (sensitivity value, 83.6%) was lower than probability of "true negatives" (specificity value, 97.1%), suggesting imperfect detection.

4 | DISCUSSION

4.1 | Comparison of quantitative and droplet digital PCR

Droplet digital PCR detected *S. spallanzanii* in a greater number of samples and had higher detection probabilities in the occupancy
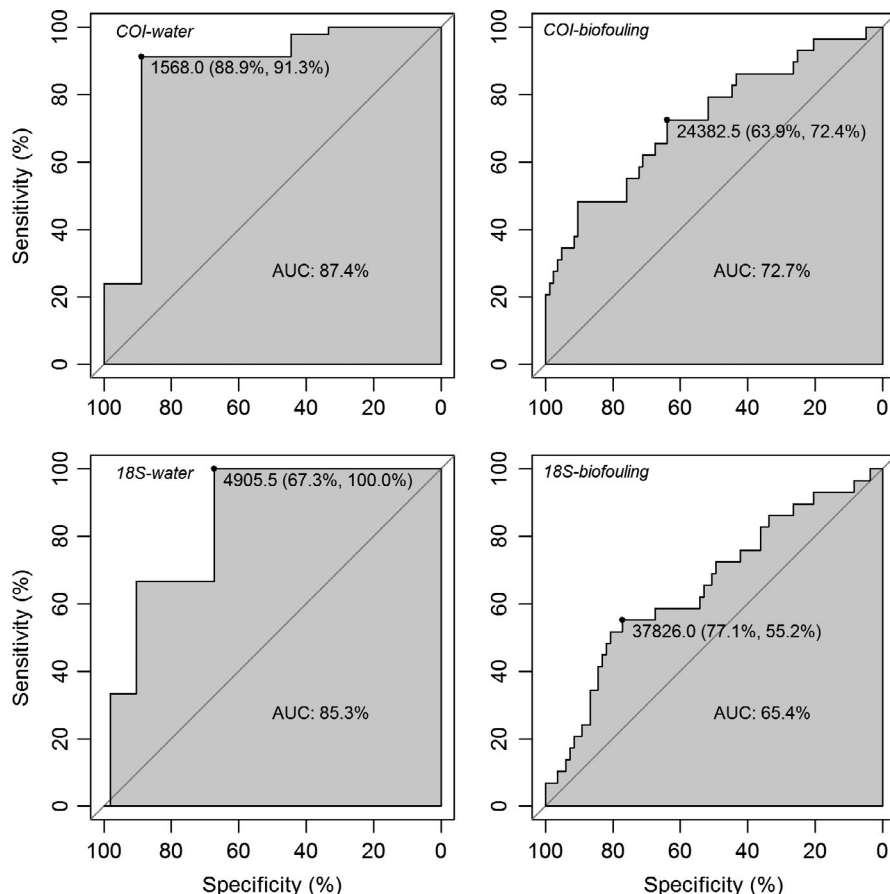
FIGURE 3 Receiver operator characteristics (ROC) with sequencing depth used as a predictor and true *Sabella spallanzanii* detections in the metabarcoding data as a binary classifier, produced for (a and b) cytochrome c oxidase I (COI) and (c and d) 18S rRNA data, water and biofouling samples, respectively. The area under curve (AUC) represents a measure of the response to the predictor, where 100% represents a perfect predictor, 50% a noninformative predictor. Values > 80% and > 70% indicate excellent and an acceptable response correspondingly. An estimate of an optimal threshold is indicated, showing the sequencing depth corresponding to the best specificity-sensitivity combination (the specificity and sensitivity values are shown in parentheses)

modelling compared to qPCR. Two likely explanations are that ddPCR is less affected by inhibitors and has greater sensitivity. Each droplet in ddPCR contains zero or one to few copies of target DNA. This distribution also occurs for inhibitors, markedly reducing their concentration (Dingle, Sedlak, Cook, & Jerome, 2013; Doi, Uchii, et al., 2015). Previously, studies have demonstrated greater sensitivity in data when comparing qPCR to ddPCR in clinical and environmental samples, e.g., ddPCR can detect differences as low as 1.25-fold, which is more accurate than the 2-fold differences detected using qPCR (Hindson et al., 2011; Nathan et al., 2014). Furthermore, smaller variations in ddPCR estimates compared to qPCR are reported, enabling more accurate quantification of eDNA concentration (Doi, Uchii, et al., 2015; Nathan et al., 2014).

There was a strong positive relationship between copy number concentrations as determined via qPCR and ddPCR, a result that has been shown previously (e.g., Kim, Jeong, & Cho, 2014; Nathan et al., 2014). However, in contrast to other studies, copy number was on average 125-fold lower when determined by qPCR. The standard curves utilized in the qPCR assay were generated from quantified amplified PCR product. This relies on accurate quantification, which is a known source of error (Bhat et al., 2010) and in future the of gBlocks, synthetically made sequence-verified, double-stranded DNA fragments, would probably reduce this source of variability. In contrast, ddPCR does not require standard curves as positive and negative droplets are separated and counted, and

the fraction of positive droplets is used to calculate target copy number.

4.2 | Comparison of Cytochrome c oxidase I and 18S rRNA metabarcoding

Numerous studies have highlighted variations in biological communities identified when different genes are targeted in high throughput sequencing (HTS) studies. This variation is commonly attributed to: lack of primer universality or specificity, preferential amplification of certain organisms (e.g., taxa with short sequences), incomplete or inaccurate reference databases, and differences in the levels of taxonomy that can be resolved by target genes (e.g., De Barba et al., 2014; Deagle et al., 2014; Hatzenbuhler, Kelly, Martinson, Okum, & Pilgrim, 2017; Wilcox et al., 2013). Although not a specific aim of this study, we observed marked differences in taxonomic composition of the communities when targeting the COI and 18S rRNA gene. These data support the need for the application of multiple gene approaches when characterizing diverse eukaryotic communities.

In the present study, the HTS data were interrogated for the presence of only one species (*S. spallanzanii*). Our in silico and previous testing with DNA from individual *S. spallanzanii* (Wood et al., 2017) showed that this species should have been amplified with the primers used, and reference sequences were present in both databases used to assign taxonomy. Despite this, *S. spallanzanii* detection rates

FIGURE 4 Receiver operator characteristics (ROC) number of sequences assigned to *Sabella spallanzani* detections in the metabarcoding data, and *S. spallanzanii* detections by either quantitative PCR or droplet digital PCR as a binary classifier, for (a and b) cytochrome c oxidase I (COI) and (c and d) 18S rRNA, water and biofouling samples, respectively. A detailed explanation of the plots is provided in Figure 3
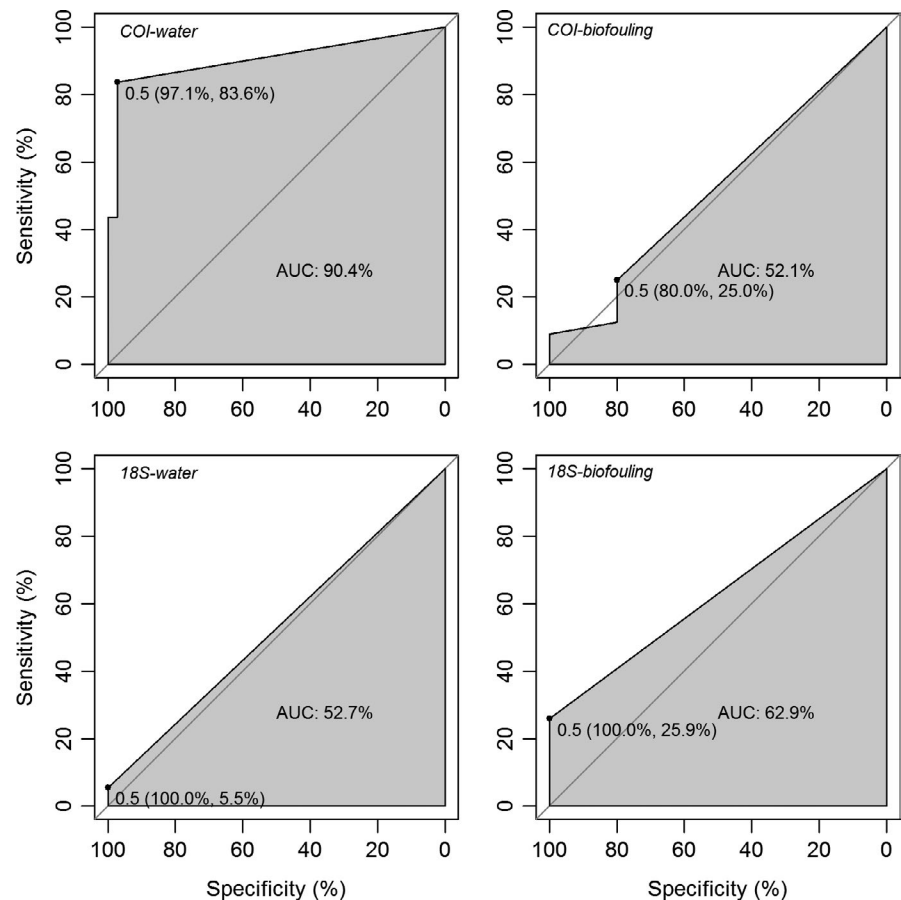
were much lower in the HTS data compared to the qPCR and ddPCR data. Two plausible explanations are: primer biases which could have a marked impact on detection of specific taxa in complex matrices, a result which has been alluded to previously in "mock community" studies (Lee at al., 2012; Pochon et al., 2013), and variability in target gene copy numbers of other taxa in the samples which differentially enhanced their amplification over that of *S. spallanzanii*. Both of the primers sets or genes targeted in this study have been used in previous marine biosecurity focused studies (e.g., von Ammon et al., 2018; Borrell, Miralles, Do Huu, Mohammed-Geba, & Garcia-Vazquez, 2017). The result of the present study highlights the need for careful consideration when using metabarcoding approaches for marine biosecurity surveillance.

Although *S. spallanzanii* detections were markedly higher for the COI gene compared to the 18S rRNA in the water samples, they were comparable in the biofouling samples. This could partially be related to the greater diversity detected in the biofouling samples when targeting the COI gene which as noted above could have resulted in preferential amplification of non-*S. spallanzanii* targets. An alternative explanation is sequencing depth. The mean reads per samples (rps) was comparable for the COI water and biofouling, and 18S rRNA biofouling (c. 31,000 rps), but markedly lower for the water 18S rRNA samples. The cause of the lower number of rps in the water samples requires further investigation.

Currently the most likely explanation is due to variability between sequencing runs, with some producing a higher number of rps than others. Metabarcoding sensitivity has been shown to improve with increase read depth per sample (Kelly, Port, Yamahara, & Crowder, 2014). The ROC analysis demonstrated that sequencing depth affects the probability of correct detection of a target species from metabarcoding data. The hypothesis that sequencing depth can predict correct *S. spallanzanii* detection was supported by high AUC values (>80%) reported for both metabarcoding markers in water samples. Notably, much higher sequencing depth is required for an optimal combination of sensitivity and specificity when detecting the species from biofouling samples. This indicates that water may be the best matrix to target during marine biosecurity surveys and suggests more stringent minimal quality criteria for metabarcoding data derived from biofouling is required. The hypothesis regarding number of putative *S. spallanzanii* sequences in metabarcoding data being a good predictor of a true detection was supported only for the COI assay applied to water samples, as evidenced by excellent response (AUC > 90%) of a binary classifier (targeted *S. spallanzanii* detection either by qPCR or ddPCR). This is probably a synergetic effect of higher detection efficiency from water samples and better taxonomic resolution, as well as more reliable taxonomic assignments attained by the COI marker (Cowart et al., 2015; Wangensteen et al., 2018) despite the cost of a higher proportion of unassigned taxa. There is a reasonable

probability of type I error (false positive detections) when using metabarcoding, as evidenced by imperfect sensitivity (83.6%), therefore, additional investigation employing targeted assays and/or visual survey couple with morphological assessments to confirm species presence is recommended.

4.3 | Targeted and community-wide characterization for detecting rare or invasive species

The occupancy modelling contributed to our understanding of detection probabilities between methodological approaches and sample types. The analysis showed that although there was a greater abundance of *S. spallanzanii* in the biofouling samples, higher probabilities of detection were derived for water samples regardless of methodology (discussed above). It also demonstrated that application of the targeted methods (qPCR or ddPCR) resulted in higher detection probabilities, with ddPCR slightly outperforming qPCR.

In the present study we report targeted detection results in copy numbers. The rational for this was to enable a comparison of the quantitative capacities and differences between qPCR and ddPCR (discussed above). However, there are some indications from work on eDNA that there can be relationships between copy numbers detected, and actual organism abundance (Chambert, Pilliod, Goldberg, Doi, & Takaharam, 2018; Dowle et al., 2016; Lacoursière-Roussel, Côté, Leclerc, & Bernatchez, 2016; Plough et al., 2018). Caution is recommended as different life stages often have variable cell numbers, DNA may be shed differentially across life stages or under varying environmental conditions and there may be differing rates of degradation. DNA has also been shown to be relatively stable, persisting in the water column for many days (Andruszkiewicz, Sassoubre, & Boehm, 2017), and therefore may have been transported into the sampling area. Despite uncertainty pertaining their ability to estimate actual population abundance, targeted approaches are extremely sensitive and specific for early detection purposes. They negate the need for morphological identification of life-stages with indistinct morphological features, and may enable larger scale surveys as they can be relatively high-throughput. The greater sensitivity of ddPCR, coupled with its high-throughput capacity, makes this method very amenable for marine biosecurity assessments when targeting one or a few species.

Despite the limitations of metabarcoding, and its lower detection of *S. spallanzanii*, it does provide a wealth of biodiversity information that could lead to a greater understanding of the impact of this invasive species on native community structure and ecosystem functioning. It is a useful tool for screening unknown communities, and although caution must be taken when raising alerts from metabarcoding data, it is useful for providing information on putative pests in parallel to targeted molecular analysis and traditional sampling approaches in marine biosecurity programmes.

5 | CONCLUSIONS

In this study we demonstrated that when applying the same assay to environmental samples (water or biofouling), ddPCR was slightly more sensitive than qPCR for the detection of *S. spallanzanii*. Both methods detected the presence of *S. spallanzanii* in a greater number of samples compared to metabarcoding targeting the COI and 18S rRNA genes. When sensitive and precise detection of one or a few species in environmental samples is paramount, targeted molecular methods are recommended. The occupancy modelling and ROC analysis indicated that the highest detection probabilities and greatest sensitivities were obtained with the water samples for all methods, possibly due to lower community complexity and/or reduced inhibitors. Also, the collection of water samples with a plankton net followed by filtration provides a relatively straightforward, time efficient and cost-effective way to sample larger volumes of initial matrix, and is a practical method that can be deployed on vessels and/or in remote field conditions.

AUTHOR CONTRIBUTIONS

S.A.W., X.P., A.Z. designed the research, S.A.W., J.A., U.A. performed the research, S.A.W., O.L., U.A., A.Z. analysed the data, S.A.W., O.L., A.Z. wrote the paper and all authors contributed to reviews.

DATA AVAILABILITY

Raw sequences from the metabarcoding are available in the National Center for Biotechnology Information Sequence Read Archive database under the accession number PRJNA477646 (water), PRJNA478269 (COI biofouling), and PRJEB25036 (18S rRNA biofouling). Quantatative PCR and ddPCR copy number and data used for occupancy modelling are available in Table S2.

ORCID

Susanna A. Wood (iD) https://orcid.org/0000-0003-1976-8266

REFERENCES

Ammon, U. V., Wood, S. A., Laroche, O., Zaiko, A., Tait, L., Lavery, S., … Pochon, X. (2018). Combining morpho-taxonomy and metabarcoding enhances the detection of non-indigenous marine pests in biofouling communities. *Scientific Reports*, 8, 16290. https://doi.org/10.1038/s41598-018-34541-1

Andruszkiewicz, E. A., Sassoubre, L. M., & Boehm, A. B. (2017). Persistence of marine fish environmental DNA and the influence of sunlight. *PLoS ONE*, 12, e0185043. https://doi.org/10.1371/journal.pone.0185043

Aronesty, E. (2011). *ea-utils*: Command-line tools for processing biological sequencing data. Retrieved from https://github.com/ExpressionAnalysis/ea-utils

Aylagas, E., Borja, Á., Irigoien, X., & Rodríguez-Ezpeleta, N. (2016). Benchmarking DNA Metabarcoding for Biodiversity-Based Monitoring and Assessment. *Frontiers in Marine Science*, 10. https://www.frontiersin.org/articles/10.3389/fmars.2016.00096/full. https://doi.org/10.3389/fmars.2016.00096

Bhat, S., Curach, N., Mostyn, T., Bains, G. S., Griffiths, K. R., & Emslie, K. R. (2010). Comparison of methods for accurate quantification of DNA mass concentration with traceability to the international system of units. *Analytical Chemistry*, 82, 7185–7192. https://doi.org/10.1021/ac100845m

Bista, I., Carvalho, G. R., Walsh, K., Seymour, M., Hajibabaei, M., Lallias, D., … Creer, S. (2017). Annual time-series analysis of aqueous eDNA reveals ecologically relevant dynamics of lake ecosystem biodiversity. *Nature Communications*, 8, 14087. https://doi.org/10.1038/ncomms14087

Borrell, Y. J., Miralles, L., Do Huu, H., Mohammed-Geba, K., & Garcia-Vazquez, E. (2017). DNA in a bottle-Rapid metabarcoding survey for early alerts of invasive species in ports. *PLoS ONE*, 12, e0183347.0.1371/journal.pone.0183347

Brown, E. A., Chain, F. J. J., Zhan, A., MacIsaac, H. J., & Cristescu, M. E. (2016). Early detection of aquatic invaders using metabarcoding reveals a high number of non-indigenous species in Canadian ports. *Diversity and Distributions*, 22, 1045–1059. https://doi.org/10.1111/ddi.12465

Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., … Wittwer, C. T. (2009). The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry*, 55, 611–622. https://doi.org/10.1373/clinchem.2008.112797

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., … Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, 7, 335–336. https://doi.org/10.1038/nmeth.f.303

Chambert, T., Pilliod, D. S., Goldberg, C. S., Doi, H., & Takaharam, T. (2018). An analytical framework for estimating aquatic species density from environmental DNA. *Ecology and Evolution*, 25, 3468–3477. https://doi.org/10.1002/ece3.3764

Clarke, L. J., Soubrier, J., Weyrich, L. S., & Cooper, A. (2014). Environmental metabarcodes for insects: in silico PCR reveals potential for taxonomic bias. *Molecular Ecology Resources*, 14, 1160–1170. https://doi.org/10.1111/1755-0998.12265

Cowart, D. A., Pinheiro, M., Mouchel, O., Maguer, M., Grall, J., Miné, J., & Arnaud-Haond, S. (2015). Metabarcoding is powerful yet still blind: A comp, arative analysis of morphological and molecular surveys of seagrass communities. *PLoS ONE*, 10, e0117562. https://doi.org/10.1371/journal.pone.0117562

Cox, M. P., Peterson, D. A., & Biggs, P. J. (2010). SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics*, 11, 485. https://doi.org/10.1186/1471-2105-11-485

Cristescu, M. E., & Hebert, P. D. N. (2018). Uses and misuses of environmental DNA in biodiversity science and conservation. *Annual Review of Ecology, Evolution, and Systematics*, 49, 209–230. https://doi.org/10.1146/annurev-ecolsys-110617-062306

De Barba, M., Miquel, C., Boyer, F., Mercier, C., Rioux, D., Coissac, E., & Taberlet, P. (2014). DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: Application to omnivorous diet. *Molecular Ecology Resources*, 14, 306–323. https://doi.org/10.1111/1755-0998.12188

de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahe, F., Logares, R., … Velayoudon, D. (2015). Eukaryotic plankton diversity in the sun-lit ocean. *Science*, 348, 1261605. https://doi.org/10.1126/science.1261605

Deagle, B. E., Jarman, S. N., Coissac, E., Pompanon, F., & Taberlet, P. (2014). DNA metabarcoding and the cytochrome c oxidase subunit I marker: Not a perfect match. *Biology Letters*, 10, 20140562. https://doi.org/10.1098/rsbl.2014.0562

Deiner, K., Bik, H. M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., … Bernatchez, L. (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology*, 26, 5872–5895. https://doi.org/10.1111/mec.14350

DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics*, 44, 837–845.

Dingle, T. C., Sedlak, R. H., Cook, L., & Jerome, K. R. (2013). Tolerance of droplet-digital PCR vs real-time quantitative PCR to inhibitory substances. *Clinical Chemistry*, 59, 1670–1672. https://doi.org/10.1373/clinchem.2013.211045

Doi, H., Takahara, T., Minamoto, T., Matsuhashi, S., Uchii, K., & Yamanaka, H. (2015). Droplet digital Polymerase Chain Reaction (PCR) outperforms real-time PCR in the detection of environmental DNA from an invasive fish species. *Environmental Science and Technology*, 49, 5601–5608. https://doi.org/10.1021/acs.est.5b00253

Doi, H., Uchii, K., Takahara, T., Matsuhashi, S., Yamanaka, H., & Minamoto, T. (2015). Use of droplet digital PCR for estimation of fish abundance and biomass in environmental DNA surveys. *PLoS ONE*, 10, e0122763. https://doi.org/10.1371/journal.pone.0122763

Dowle, E., Pochon, X., Banks, J., Shearer, K., & Wood, S. A. (2016). Targeted gene enrichment and high throughput sequencing for environmental biomonitoring. *Molecular Ecology Resources*, 16, 1240–1254. https://doi.org/10.1111/1755-0998

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26, 2460–2461. https://doi.org/10.1093/bioinformatics/btq461

Evans, N. T., Li, Y., Renshaw, M. A., Olds, B. P., Deiner, K., Turner, C. R., Pfrender, M. E. (2017). Fish community assessment with eDNA metabarcoding: Effects of sampling design and bioinformatic filtering. *Canadian Journal of Fisheries and Aquatic Sciences*, 74(9), 1362–1374. https://doi.org/10.1139/cjfas-2016-0306

Gillum, J. E., Jimenez, L., White, D. J., Goldstien, S. J., & Gemmell, N. J. (2014). Development and application of a quantitative real-time PCR assay for the globally invasive tunicate Styela clava. *Management of Biological Invasions*, 5, 133–142.

Goldberg, C. S., Turner, C. R., Deiner, K., Klymus, K. E., Philip, K., Thomsen, F., … Taberlet, P. (2016). Critical consideration for the application of environmental DNA methods to detect aquatic species. *Methods in Ecology and Evolution*, 7, 1299–1307. https://doi.org/10.1111/2041-210X.12595

Guillera-Arroita, G., Ridout, M. S., & Morgan, B. J. T. (2010). Design of occupancy studies with imperfect detection. *Methods in Ecology and Evolution*, 1, 131–139. https://doi.org/10.1111/j.2041-210X.2010.00017

Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., … Christen, R. (2013). The Protist Ribosomal Reference database (PR2): A catalog of unicellular eukaryote small sub-unit rRNA sequences

with curated taxonomy. *Nucleic Acids Research*, 41, D59–604. https://doi.org/10.1093/nar/gks1160

Hale, S. S., & Heltche, J. F. (2008). Signals from the benthos: Development and evaluation of a benthic index for the nearshore Gulf of Maine. *Ecological Indicators*, 8, 338–350. https://doi.org/10.1016/j.ecolind.2007.04.004

Hänfling, B., Lawson Handley, L., Read, D. S., Hahn, C., Li, J., Nichols, P., ... Winfield, I. J. (2016). Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods. *Molecular Ecology*, 25, 3101–3119. https://doi.org/10.1111/mec.13660

Hatzenbuhler, C., Kelly, J. R., Martinson, J., Okum, S., & Pilgrim, E. (2017). Sensitivity and accuracy of high-throughput metabarcoding methods for early detection of invasive fish species. *Scientific Reports*, 13, 46393. https://doi.org/10.1038/srep46393

Haugland, R. A., Siefring, S. C., Wymer, L. J., Brenner, K. P., & Dufour, A. P. (2005). Comparison of *Enterococcus* measurements in freshwater at two recreational beaches by quantitative polymerase chain reaction and membrane filter culture analysis. *Water Research*, 39, 559–568. https://doi.org/10.1016/j.watres.2004.11.011

Hindson, B. J., Ness, K. D., Masquelier, D. A., Belgrader, P., Heredia, N. J., Makarewicz, A. J., ... Colston, B. W. (2011). High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Analytical Chemistry*, 83, 8604–8610. https://doi.org/10.1021/ac202028g

Hines, J. E. (2006). PRESENCE – Software to estimate patch occupancy and related parameters. USGS-PWRC. Retrieved from http://www.mbr-pwrc.usgs.gov/software/presence.html.

Holloway, M. G., & Keough, M. J. (2002). An introduced polychaete affects recruitment and larval abundance of sessile invertebrates. *Ecological Application*, 12, 1803–1823. https://doi.org/10.1890/1051-0761

Keeley, N., Wood, S. A., & Pochon, X. (2018). Development and validation of a multi-trophic metabarcoding biotic index for benthic organic enrichment biomonitoring. *Ecological Indicators*, 85, 1044–1057. https://doi.org/10.1016/j.ecolind.2017.11.014

Kelly, R. P., Port, J. A., Yamahara, K. M., & Crowder, L. B. (2014). Using environmental DNA to census marine fishes in a large mesocosm. *PLoS ONE*, 9, e86175. https://doi.org/10.1371/journal.pone.0086175.

Kim, T. G., Jeong, S. Y., & Cho, K. S. (2014). Comparison of droplet digital PCR and quantitative real-time PCR for examining population dynamics of bacteria in soil. *Applied Microbiology Biotechnology*, 98, 6105–6113. https://doi.org/10.1007/s00253-014-5794-4

Kreader, C. A. (1996). Relief of amplification inhibition in PCR with bovine serum albumin or T4 gene 32 protein. *Applied and Environmental Microbiology*, 62, 1102–1106.

Lacoursière-Roussel, A., Côté, G., Leclerc, V., & Bernatchez, L. (2016). Quantifying relative fish abundance with eDNA: A promising tool for fisheries management. *Journal of Applied Ecology*, 53, 1148–1157. https://doi.org/10.1111/1365-2664.12598

Laroche, O., Wood, S. A., Tremblay, L. A., Ellis, J. I., Lejzerowicz, F., Pawlowski, J., ... Pochon, X. (2016). First evaluation of foraminiferal metabarcoding for monitoring benthic health at an offshore oil drilling site. *Marine Environmental Research*, 120, 225–235. https://doi.org/10.1016/j.marenvres.2016.08.009

Lee, C. K., Herbold, C. W., Polson, S. W., Wommack, K. E., Williamson, S. J., McDonald, I. R., & Cary, S. C. (2012). Groundtruthing next-gen sequencing for microbial ecology-biases and errors in community structure estimates from PCR amplicon pyrosequencing. *PLoS ONE*, 7(9), e44224. https://doi.org/10.1371/journal.pone.0044224

Leray, M., Yang, J. Y., Meyer, C. P., Mills, S. C., Agudelo, N., Ranwez, V., ... Machida, R. J. (2013). A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: Application for characterizing coral reef fish gut contents. *Frontiers in Zoology*, 10, 14. https://doi.org/10.1186/1742-9994-10-34

Lugg, W. H., Griffiths, J., van Rooyen, A. R., Weeks, A. R., & Tingley, R. (2017). Optimal survey designs for environmental DNA sampling. *Methods in Ecology and Evolution*, 9(4), 1049–1059. https://doi.org/10.1111/2041-210X.12951

Machida, R. J., Leray, M., Ho, S.-L., & Knowlton, N. (2017). Metazoan mitochondrial gene sequence reference datasets for taxonomic assignment of environmental samples. *Scientific Data*, 4, 1–7. https://doi.org/10.1038/sdata.2017.27

MacKenzie, D. I., Nichols, J. D., Lachman, G. D., Droege, S., Royle, J. A., & Langtimm, C. A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, 83, 2248–2255. https://doi.org/10.1890/0012-9658(2002)083[2248:ESORWD]2.0.CO;2

Mahé, F., Rognes, T., Quince, C., de Vargas, C., & Dunthorn, M. (2014). Swarm: Robust and fast clustering method for amplicon-based studies. *PeerJ*, 2, e593. https://doi.org/10.7717/peerj.593

Murtaugh, P. A. (1996). The statistical evaluation of ecological indicators. *Ecological Applications*, 6, 132–139. https://doi.org/10.2307/2269559

Nathan, L. R., Simmons, M. D., Wegleitner, B., Jerde, C. L., & Mahon, A. (2014). Quantifying environmental DNA signals for aquatic invasive species across multiple detection platforms. *Environmental and Science Technology*, 48, 12800–12806. https://doi.org/10.1021/es5034052

Nichols, J. D., Bailey, L. L., O'Connell, A. F. Jr, Talancy, N. W., Campbell Grant, E. H., Gilbert, A. T., ... Hines, J. E. (2008). Multi-scale occupancy estimation and modeling using multiple detection methods. *Journal of Applied Ecology*, 45, 1321–1329. https://doi.org/10.1111/j.1365-2664.2008.01509.x

Nilsson, R. H., Ryberg, M., Kristiansson, E., Abarenkov, K., Larsson, K. H., & Kõljalg, U. (2006). Taxonomic reliability of DNA sequences in public sequence databases: A fungal perspective. *PLoS ONE*, 1, e59. https://doi.org/10.1371/journal.pone.0000059

Oliveros, J. C. (2015) Venny. An interactive tool for comparing lists with Venn's diagrams. Retrieved from http://bioinfogp.cnb.csic.es/tools/venny/index.html

Opel, K. L., Chung, D., & McCord, B. R. (2010). A study of PCR inhabitation mechanisms using real time PCR. *Journal of Forensic Science*, 55, 25–33. https://doi.org/10.1111/j.1556-4029.2009.01245.x

Patti, F. P., & Gambi, M. C. (2001). Phylogeography of the invasive polychaete *Sabella spallanzanii* (Sabellidae) based on the nucleotide sequence of internal transcribed spacer 2 (ITS2) of nuclear rDNA. *Marine Ecology Progress Series*, 215, 169–177. https://doi.org/10.1017/S0025315417000261P

Plough, L. V., Ogburn, M. B., Fitzgerald, C. L., Geranio, R., Marafino, G. A., & Richie, K. D. (2018). Environmental DNA analysis of river herring in Chesapeake Bay: A powerful tool for monitoring threatened keystone species. *PLoS ONE*, 13, e0205578. https://doi.org/10.1371/journal.pone.0205578

Pochon, X., Bott, N., Smith, K. F., & Wood, S. A. (2013). Evaluating detection limits of Next Generation Sequencing for the surveillance and monitoring of international marine pests. *PLoS ONE*, 8, e73935. https://doi.org/10.1371/journal.pone.0073935

Pochon, X., Zaiko, A., Fletcher, L. M., Laroche, O., & Wood, S. A. (2017). Wanted dead or alive? Using metabarcoding of environmental DNA and RNA to distinguish living assemblages for biosecurity applications. *PLoS ONE*, 12, e0187636. https://doi.org/10.1371/journal.pone.0187636

Pregler, K. C., Vokoun, J. C., Jensen, T., & Hagstrom, N. (2015). Using multimethod occupancy estimation models to quantify gear differences in detection probabilities: Is backpack electrofishing missing occurences for a species of concern? *Transactions of the American Fishing Society*, 141, 89–95. https://doi.org/10.1080/00028487.2014.968291

R Development Core Team (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org

Ratnasingham, S., & Hebert, P. D. N. (2007). BOLD: The barcode of life data system (www.barcodinglife.org). *Molecular Ecology Notes*, 7, 355–364. https://doi.org/10.1111/j.1471-8286.2006.01678.x

Read, G. B., Inglis, G. J., Stratford, P., & Ahyong, S. T. (2011). Arrival of the alien fanworm *Sabella spallanzanii* (Gmelin, 1791) (Polychaeta: Sabellidae) in two New Zealand harbours. *Aquatic Invasion*, 6, 273–279. https://doi.org/10.3391/ai.2011.6.3.04

Robin, X., Turk, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, 12, 77. https://doi.org/10.1186/1471-2105-12-77

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: A versatile open source tool for metagenomics. *PeerJ*, 4, e2584. https://doi.org/10.7717/peerj.2584

Sigsgaard, E. E., Carl, H., Møller, P. R., & Thomsen, P. F. (2015). Monitoring the near-extinct European weather loach in Denmark based on environmental DNA from water samples. *Biological Conservation*, 183, 46–52. https://doi.org/10.1016/j.biocon.2014.11.023

Smith, K. F., Wood, S. A., Mountfort, D., & Cary, S. C. (2012). Development of a real-time PCR assay for the detection of the invasive clam, *Corbula amurensis*, in environmental samples. *Journal of Experimental and Marine Biology and Ecology*, 412, 52–57. https://doi.org/10.1016/j.jembe.2011.10.021

Taberlet, P., Bonin, A., Zinger, L., & Coissac, E. (Eds) (2018). *Environmental DNA for biodiversity research and monitoring*. Oxford, UK: Oxford University Press.

Tait, L., Inglis, G., & Seaward, K. (2018). Enhancing passive sampling tools for detecting marine bioinvasions. *Marine Pollution Bulletin*, 128, 41–50. https://doi.org/10.1016/j.marpolbul.2018.01.015

Takahara, T., Minamoto, T., & Doi, H. (2013). Using environmental DNA to estimate the distribution of an invasive fish species in ponds. *PLoS ONE*, 8, e56584. https://doi.org/10.1371/journal.pone.0056584

Te, S. H., Chen, E. Y., & Gin, K. Y. (2015). Comparison of quantitative PCR and droplet digital PCR multiplex assays for two genera of bloom-forming cyanobacteria, *Cylindrospermopsis* and *Microcystis*. *Applied and Environmental Microbiology*, 81(5203–5211), 5203–5211. https://doi.org/10.1128/AEM.00931-15

Thomsen, P. F., Kielgast, J., Iversen, L. L., Møller, P. R., Rasmussen, M., & Willerslev, E. (2012). Detection of a diverse marine fish fauna using environmental DNA from seawater samples. *PLoS ONE*, 7, e41732. https://doi.org/10.1371/journal.pone.0041732

von Ammon, U., Wood, S. A., Laroche, O., Zaiko, A., Tait, L., Lavery, S., … Pochon, X. (2017). The impact of artificial surfaces on marine bacterial and eukaryotic biofouling assemblages: A high-throughput sequencing analysis. *Marine Environmental Research*, 133, 57–66. https://doi.org/10.1016/j.marenvres.2017.12.003

Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naive bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, 73, 5261–5267. https://doi.org/10.1128/AEM.00062-07

Wangensteen, O. S., Palacín, C., Guardiola, M., & Turon, X. (2018). DNA metabarcoding of littoral hard-bottom communities: High diversity and database gaps revealed by two molecular markers. *PeerJ*, 6, e4705. https://doi.org/10.7717/peerj.4705

Wilcox, T. M., McKelvey, K. S., Young, M. K., Jane, S. F., Lowe, W. H., Whiteley, A. R., & Schwartz, M. K. (2013). Robust detection of rare species using environmental DNA: The importance of primer specificity. *PLoS ONE*, 8(3), e59520. https://doi.org/10.1371/journal.pone.0059520

Wood, S. A., Pochon, X., Ming, W., von Ammn, U., Woods, C., Carter, M., … Zaiko, A. (2018). Considerations for incorporating real-time PCR assays into routine marine biosecurity surveillance programmes: A case study targeting the Mediterranean fanworm (*Sabella spallanzanii*) and club tunicate (*Styela clava*). *Genome*, 62, 137–146. https://doi.org/10.1139/gen-2018-0021

Wood, S. A., Smith, K. F., Banks, J. C., Tremblay, L., Rhodes, L., Mountfort, D., … & Pochon, X. (2013). Molecular genetic tools for environmental monitoring of New Zealand's aquatic habitats, past, present and the future. *New Zealand Journal of Marine and Freshwater Research*, 47, 90–119.

Wood, S. A., Zaiko, A., Richter, I., Inglis, G., & Pochon, X. (2017). Development of a real-time Polymerase Chain Reaction assay for the detection of the invasive Mediterranean fanworm, *Sabella spallanzanii*, in environmental samples. *Environmental Science and Pollution Research*, 24, 17373–17382. https://doi.org/10.1007/s11356-017-9357-y

Zaiko, A., Schimanski, K., Pochon, X., Hopkins, G. A., Goldstien, S., Floerl, O., & Wood, S. A. (2016). Metabarcoding improves detection of eukaryotes from early biofouling communities: Implications for pest monitoring and pathway management. *Biofouling*, 32, 671–684. https://doi.org/10.1080/08927014.2016.1186165

Zhan, A., Hulák, M., Sylvester, F., Huang, X., Adebayo, A. A., Abbott, C. L., … MacIsaac, H. J. (2013). High sensitivity of 454 pyrosequencing for detection of rare species in aquatic communities. *Methods in Ecology and Evolution*, 4, 558–565. https://doi.org/10.1111/2041-210X.12037

Zinger, L., Bonin, A., Alsos, I. G., Bálint, M., Bik, H., Boyer, F., … Taberlet, P. (2019). DNA metabarcoding—Need for robust experimental designs to draw sound ecological conclusions. *Molecular Ecology*, 28(8), 1857–1862. https://doi.org/10.1111/mec.15060

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.