GUIDELINES FOR DEVELOPMENT
and validation of assays for marine pests

Australian Government
Department of Agriculture
and Water Resources

Cataloguing data
This publication (and any material sourced from it) should be attributed as: Department of Agriculture and Water Resources 2018, Guidelines for the development and validation of assays for marine pests, Department of Agriculture and Water Resources, Canberra, May. CC BY 4.0.

ISBN 978-1-76003-156-5
This publication is available at marinepests.gov.au
Department of Agriculture and Water Resources
GPO Box 858 Canberra ACT 2601
Telephone 1800 900 090
Web agriculture.gov.au

Acknowledgements
Guidelines for development and validation of assays for marine pests was developed based on a workshop that was held in Adelaide 16 November 2016. Workshop members included:

Sarah Graham	Australian Government Department of Agriculture and Water Resources
Ingo Ernst	Australian Government Department of Agriculture and Water Resources
Joanne Banks	Australian Government Department of Agriculture and Water Resources
William Wong	Australian Government Department of Agriculture and Water Resources
Peter Stoutjesdijk	Australian Government Department of Agriculture and Water Resources
Nickolas Fahy	Australian Government Department of Agriculture and Water Resources
Tony Arthur	Australian Government Department of Agriculture and Water Resources
Jeffrey Go	New South Wales Department of Primary Industries
Cheryl Jenkins	New South Wales Department of Primary Industries
Anita Ramage	Queensland Department of Agriculture and Fisheries
Kevin Ellard	Tasmania Department of Primary Industries, Parks, Water and Environment
Murray Barton	Northern Territory Department of Primary Industry and Resources
Michelle Besley	South Australia Department of Primary Industries and Regions
Marty Deveney	South Australia Department of Primary Industries and Regions
Kathryn Wiltshire	South Australia Department of Primary Industries and Regions
Susie Wood	New Zealand Cawthron Institute
Nick Moody	Australian Animal Health Laboratory, Commonwealth Scientific and Industrial Research Organisation
Elise Furlan	University of Canberra
Michael Bunce	Curtin University
Roger Huerlimann	James Cook University
Andrew Weeks	University of Melbourne
Evan Sergeant	AusVet

Contributions to the development of this publication are gratefully acknowledged.

Cover image sources: Diver – Ægir Divers / Ship image - Gordon Bell / Northern Pacific seastar - Justin McDonald

Table of Contents

Introduction

Marine pests have the potential to significantly affect marine industries and environments around the world. Marine pests continue to spread progressively through international and domestic vessel movements (such as ballast water and biofouling) and other vectors (such as ghost fishing nets and flotsam). Efforts to prevent, control or eradicate pests are challenged by difficulties in early detection and understanding of pest abundance and distribution in the environment. Traditional methods of detection are expensive and labour intensive, relying on morphological identification through visual surveys, or sample collection by traps, trawls and other means. Developments in environmental sampling and molecular technology have led to the concurrent development of new methods for the detection of marine pest DNA in environmental samples and/or DNA-based identification of specific pests.

To maximise the utility of new assays applied for pest detection or identification, it is important that assay performance is predictable and well understood, to support interpretation of assay results and consequent decision-making. Any new assay development must meet agreed national guidelines for assay development and validation, to ensure that the resulting assay will provide a useful tool for future surveillance activities.

For terrestrial and aquatic animal disease assays, there is an internationally recognised assay validation process promulgated by the World Organisation for Animal Health (OIE). However, despite the apparent need, there are no such guidelines available for marine pest assays. These guidelines were developed to meet that need.

These guidelines provide a framework for assay development and validation for new molecular methods for the detection or identification of marine pests. The guidelines follow a similar structure to the OIE guidelines for validation of diagnostic assays for infectious diseases (Chapter 1.1.2 of the OIE *Aquatic Manual* [OIE 2016] and Chapters 1.1.6 and 3.6 of the *Terrestrial Manual* [OIE 2016a]). The OIE guidelines were modified to suit the unique requirements of marine pest assays following discussion among professionals working in this area at a workshop in Adelaide in November 2016. Readers are referred to the OIE *Aquatic* and *Terrestrial Manuals* for more detailed information where necessary.

It is also important to note that while the development and validation pathways are described in a sequential fashion, in practice many of the steps can be undertaken in parallel and some steps may be iterative, requiring repeated attempts to achieve the desired assay performance.

1 Preliminary considerations in assay development and validation

A variety of factors that could have an impact on assay performance may affect the development and validation process. Factors to consider should include:

1.1 Collaborative relationships

Developers should, where possible, establish collaborative relationships with other institutions early in the development process, to make use of specialist skills and laboratories in testing repeatability of assays. Such approaches will facilitate the development process and support ongoing implementation of the assay and comparison/standardisation of assays across multiple laboratories with different platforms and processes.

1.2 Quality systems

Laboratories should have in place appropriate (fit for purpose) quality assurance and management systems, to minimise the influence of factors that would negatively the consistency and quality of assay results. The use of validated tests with positive and negative controls and an appreciation of the sensitivity and specificity of the test must be considered during test development and validation. Quality systems should address issues such as:

- instrumentation
- operator error
- reagent choice (both chemical and biological) and calibration
- reaction vessels and platforms
- water quality
- pH ionicity of buffers and diluents
- incubation temperatures and durations
- errors in the technical performance of the assay.

1.3 Factors affecting assay performance

Factors affecting assay performance are a critical starting point in assay development as they will affect the development process and the assay's ability to meet defined validation criteria (see below). Factors affecting assay performance can be broadly grouped as:

- sample related
  - individual or pooled samples
  - sample type
  - matrix composition and volume
  - target analyte quantity and/or quality
- the assay system

  - physical, chemical, biological and operator-related influences on the capacity of the assay to detect a specific analyte in the sample

- test result interpretation
  - the capacity of a test result to accurately predict the pest status of the source environment, considering the purpose for which the assay is applied.

Specific issues for consideration include (but are not limited to):

- selection, collection, preparation, preservation and management of samples as well as factors that decrease sample quality
- sample transport, chain of custody, sample storage and the laboratory information management system
- factors that decrease sample quality and integrity throughout the assay process
- reference samples used in further assay development and validation (For initial assay design and testing, pure seawater may be used as the substrate, but for further assay development and validation, reference samples should be in the same matrix(es) that are to be used regularly throughout the assay.)
- reference materials that appropriately represent the range of known analyte concentrations to be detected by the assay
- possible presence of endogenous or exogenous inhibitors or competitors in the sample matrix, including:
  - inhibitors and other factors affecting polymerase chain reaction (PCR) test performance
  - contamination or deterioration of the sample or
  - non-target species that affect test performance.

1.4 Target species

Pests of national priority should be considered in determining appropriate initial target species for assay development. The Consultative Committee on Introduced Marine Pest Emergencies (CCIMPE) Trigger List, lists species that are considered nationally important. The Australian Priority Marine Pest List is being developed and pests identified on the list will be agreed upon by the Australian, state and Northern Territory governments as being nationally significant for purposes of cost sharing and awareness. They are therefore likely to receive stronger support for development and validation.

2 Criteria for assay development and validation

Assay validation criteria are those attributes of an assay that provide confidence in assay results and support the use of an assay for the purpose(s) for which it has been developed. An assay should be evaluated against the following validation criteria to ensure that it is fit for the proposed purpose (see Manual of Diagnostic Tests for Aquatic Animals):

- definition of the intended purpose(s)
- optimisation
- standardisation
- repeatability
- analytical sensitivity
- analytical specificity
- test interpretation criteria
- detection sensitivity
- detection specificity
- reproducibility
- fitness for use
- suitability of multiplex tests (to detect related organisms).

The assay development and validation process can be further divided into three primary steps:

- assay development, including experimental and laboratory-based studies for proof of concept, optimisation and standardisation and initial evaluation of assay operating characteristics
- assay validation, including determining the fitness of the assay for its intended purpose(s)
- monitoring performance and maintenance of validation, including providing assurance that the validated assay consistently maintains the previously defined performance characteristics.

3 Definition of intended purpose for an assay

The first step in assay development is to define the purpose of the assay, because this guides all subsequent steps in the development and validation process.

The purpose(s) for marine pest assays generally fall into several broad categories, including:

- *Early detection of incursions:* early detection of diseases generally relies on passive surveillance. Reliance on passive surveillance alone for early detection of marine pests may not be an adequate or suitable approach for species that are hard to detect or cryptic (for example *Didemnum* sp). Active surveillance to detect marine pests needs to be cost-effective and have a high detection sensitivity. Early detection also informs the determination of port status for regulatory purposes.
- *Pest freedom:* sampling an environment to determine if a target pest is present or absent. This requires a defensible, highly sensitive and specific test to provide confidence that results are an accurate reflection of the true state of nature. It must include the determination of assay cutoff points to allow for the presence of DNA fragments or non-viable organisms.
- *Confirmation of species identity:* requires high analytical and detection specificity, particularly where there are multiple closely related species that require differentiation.
- *Abundance:* monitoring changes in pest abundance in an environment over time requires high detection sensitivity and specificity to enable interpretation of results.
- *Provenance (molecular epidemiology):* tracing the source of marine pest incursions requires high analytical and detection specificity.
- *Ballast water management:* tests must be cost-effective and have a high detection sensitivity for International Maritime Organization purposes.

This list should not be considered definitive and other less common purposes could be considered. Assay 'purposes' may also be context and location-specific.

4 Assay development

The assay development phase is primarily focused on experimental work to develop, standardise and optimise a working assay, ready for validation.

Key considerations for assay development include:

- Possible variations between laboratories (for example, assay platforms, equipment or reagents, source of positive and negative control material to support validation) should be taken into account during the development process but should provide comparable end results.
- Development of proficiency programmes to support alignment of inter-laboratory standards for assays to be used nationally.
- Assay target(s). Is the assay planned to be single-species, multi-species, genus-specific or specific at some other level?
- Type of test (for example conventional versus real-time PCR and the use of multiplex assays).
- Workflow and infrastructure required for assay implementation and quality control.
- Budget and funding for assay development.
- Availability of a library of related organisms (different genotypes, different regions, similar species) to aid in understanding inter-specific and intra-specific diversity.
- Assay optimisation and calibration issues.
- An understanding of sequence variation.
- Establishment of a collaborative approach for assay development. This may facilitate the development process, ongoing implementation and comparison between multiple laboratories.

4.1 Assay method design and proof of concept

Careful planning and design of a new assay is critical, to ensure that the assay performs to expectations and is fit for the stated purpose(s). Specific considerations include the development of analyte reference samples for use during the development and validation processes. Samples for initial assay development should contain DNA extracted from a reliably identified organism. Reference samples for validation should reflect the target analyte, the matrix in which the analyte is found, and the environment in which the assay is intended to be used. Sequence analysis for the target DNA, close taxonomic organisms and design of suitable probes for the assay are also important in assay planning and design.

Further information is available in Chapter 3.6 of the OIE *Terrestrial Manual* (OIE 2016a).

4.2 Operating range of the assay

The operating range of an assay is the range of analyte concentrations over which the assay method provides suitable *accuracy* and *precision*. Accuracy is a measure of the closeness of a test value to the expected (true) value (mean or median) for a reference standard reagent of known concentration, often estimated as the detection sensitivity and specificity of the assay. Precision is the degree of dispersion (variance, standard deviation or coefficient of variation) within a series of measurements

of the same sample tested under specified conditions, usually estimated as repeatability and reproducibility of the assay.

To formally determine this range, a high positive reference sample is selected. This high positive sample is serially diluted to extinction of the assay's response in an analyte-negative matrix of the same constitution as the sample matrix. The results may be plotted as a 'response-curve' to establish the working range of the assay. If the range is found to be unacceptable for the intended purpose, additional optimisation may be needed.

This process should be repeated with DNA in a pure matrix, with DNA in the target matrix, then with the target analyte (the organism or a reasonable facsimile) in the target matrix.

4.3 Optimisation and standardisation

Optimisation is the process of evaluation and adjustment of the physical, chemical and biological parameters of an assay to ensure that the performance characteristics of the assay are appropriate for its intended use. Optimising an assay is essential to ensure a reliable and predictable assay performance. Scientific judgement and use of best scientific practices are recommended to guide optimisation of all elements of assay development and validation.

For molecular assays, optimisation requires:

- consistency of suppliers, reagents and primer/probe design software to minimise variation
- samples with target at the limits of detection
- consistent workflows to make optimisation and standardisation easier
- separate evaluation of different sample types
- preparation of well-defined reference samples representative of the sample type to be used subsequently in the assay.

Ideally, reference samples should range from high positive to negative (for example high and low positive and negative) analyte concentrations and be sourced from the environment in which the assay is to be applied. However, in many cases (such as exotic pest assays) spiking a sample matrix with a known amount of the analyte may be the only option for creation of reference samples. The matrix, into which the analyte is placed or diluted, should be identical to, or resemble as closely as possible, the samples that will ultimately be tested in the assay. These samples are used in experiments to determine if the assay can distinguish between varying quantities of analyte, distinguish the target from closely related analytes, and to optimise reagent concentrations and perfect the protocol. Sufficient amounts of each reference sample should be prepared and stored for use throughout the entire development and validation process, as changing reference samples during the validation process is highly undesirable and adversely effects the integrity of the development and validation process.

Further information on optimisation and standardisation is provided in Chapters 1.1.6 and 3.6 of the *Terrestrial Manual* (OIE 2016a).

4.4 The sample matrix

The sampling unit, sample matrixes and volumes to be used in the assay must be clearly defined and documented. The validation process undertaken is only applicable to the range of sample types, matrixes, collection methods and volumes that are assessed during the validation process.

Validated assay performance characteristics are not applicable if the assay is used on sample types that have not been included in the validation process, or where equivalency has not been demonstrated.

Different matrixes to be used in an assay should all be evaluated during the development and validation process. Some sample matrixes can include inhibitory factors that interfere with assay performance. The occurrence of inhibitory factors should be identified and methods to overcome the inhibition developed. Where inhibition cannot be reliably overcome, it may be necessary to exclude that sample matrix from routine use in the assay

4.5 Robustness

Robustness refers to an assay's capacity to remain unaffected by minor variations in test situations that may occur over the course of testing in a single laboratory. Assessment of robustness should begin during assay development and optimisation stages. The deliberate variations in method parameters may be addressed in experiments after optimal conditions for an assay are established.

The factors most likely to affect assay robustness include pH, instrumentation, temperature, batch of reagents or brand of microtitre plates and aqueous or organic matrix factors. Once optimisation is complete, the robustness of the assay becomes part of the assessment of repeatability.

4.6 Calibration of the assay to standard reagents

4.6.1 International and national reference standards

Where possible, new assays should be calibrated to national or international reference standards, containing a known concentration of analyte. However, for exotic marine pests there are often no such standards available and this may not be possible.

4.6.2 In-house standard

In the absence of national or international reference standards, assays should be calibrated to a well-characterised in-house standard. Ideally, non-viable type material or stabilised extract should be made available for positive controls.

4.6.3 Working standard

One or more working standards should be calibrated to an international, national, or in-house standard, and are prepared in large quantities, aliquoted and stored for routine use as positive controls in each run of the assay. If the test is to be used in a number of laboratories, sharing of aliquoted standards should be considered.

4.7 'Normalising' test results to a working standard

To allow comparability of test results both within and between laboratories, one or more working standard reference samples should be included in each run of an assay. Raw result values for each test run sample can then be converted to units of activity relative to the working standard(s) by a

process called 'normalisation'. The resulting normalised results are thereby adjusted for variations in assay performance between runs and between laboratories, allowing comparability of results.

4.8 Preliminary study of repeatability

Repeatability is the level of agreement between results of replicates of a sample both within and between runs of the same test method in a given laboratory. Assessment of repeatability should begin during assay development and optimisation stages. Early indications of poor repeatability should trigger a review of methodology to improve repeatability, or a decision to abandon validation of the assay if the poor performance cannot be overcome.

Repeatability is further verified during assay validation (see below) and on an ongoing basis when the optimised test is run under routine laboratory or field conditions, as part of process control procedures for the duration of the life of the assay.

Where ever possible, validated tests should be accredited under national standards (for example the National Association of Testing Authorities, Australia).

5 Assay validation pathway

Validation is the process of evaluating the analytical and detection performance characteristics of a test, to demonstrate that it has been properly developed, optimised and standardised for an intended purpose(s). The validation process includes steps to evaluate analytical sensitivity and specificity, detection sensitivity and specificity, and reproducibility and repeatability of the assay. When assessing analytical performance, it is also important to consider available assay platforms for quantitative PCR and/or genome sequencing and ensure that performance is evaluated for consistency across the range of platforms where it is likely to be implemented. Assay validation may also be required for these different platforms and methods.

5.1 Analytical performance characteristics

Analytical performance of an assay is evaluated through experimental studies to define the ability of the assay to reliably detect the target analyte (analytical sensitivity) with a high level of precision (repeatability) and no cross-reaction with non-target species (analytical specificity).

5.1.1 Repeatability

Repeatability is evaluated by estimating the variability in results of multiple replicates of each sample and is often expressed as the *coefficient of variation* for assays that produce a quantitative result. Usually, multiple samples representing the range of target concentrations expected to be analysed should be included in the evaluation. The number of replicates should be adequate to provide valid estimates of the coefficient of variation and may be determined in consultation with a statistician.

Between-run variation is determined by using the same samples in multiple runs involving a combination of different operators over multiple days.

When evaluating repeatability, it is generally not acceptable to prepare a final working dilution of a sample in a single tube from which diluted aliquots are pipetted into reaction vessels, nor to create replicates from one extraction of nucleic acid. Rather, extract each replicate before dilution into the reaction vessels. There may be exceptions depending on the requirements of the test.

5.1.2 Analytical specificity

Analytical specificity is a qualitative assessment of the ability of an assay to differentiate the target analyte (organism) from other non-target organisms or analytes that may be present in a sample matrix. This is usually evaluated by assessing assay cross-reactivity against a panel of known organisms that are usually either closely related or known/suspected to produce analytes that are likely to be cross-reactive (for example similar genetic components). The choice and sources of sample types, organisms and sequences for the analytical specificity evaluation should reflect test purpose and assay type.

Specific issues to consider in evaluating analytical specificity for marine pest assays include:

- A library of related organisms and sequences will be required for inclusion in the assay evaluation.
- Measures to prevent laboratory contamination of samples should be identified and implemented.

- Analytical specificity should be evaluated in a range of sample matrixes that are likely to be presented once the test is implemented.
- Lack of analytical specificity should trigger investigation of the cross-reacting organisms and redesign of the assay to eliminate cross-reaction.

5.1.3 Analytical sensitivity

Analytical sensitivity of an assay is usually estimated as the limit of detection (LOD) of the target analyte. The LOD is the minimum concentration of target analyte in the sample matrix that produces a positive result a specified percentage of the time. The LOD is usually estimated by spiking a known concentration of analyte into a sample matrix and testing serial dilutions of the spiked sample until the assay produces consistently negative results.

Specific issues to consider in evaluating analytical specificity for marine pest assays include:

- analytical sensitivity should be evaluated in a range of sample matrixes that are likely to be presented before the test is implemented
- the effect of different sample volumes, filtration and processing methods on sensitivity should be evaluated
- issues associated with DNA concentration and low copy-number target DNA in the matrix (trace level detection) should be assessed.

5.2 Detection performance of the assay

Detection performance of an assay is a prime consideration in determining its fitness for purpose. Sensitivity and specificity of an assay and the relative importance of false negatives (poor sensitivity) and false positives (poor specificity) are critical to determining an assay's ability to effectively meet the proposed purpose. In general:

- assays should be target-specific, have a complementary assay which is specific, or be able to differentiate species
- false positives are likely to result in more onerous requirements and increased cost for follow-up investigation and confirmation, particularly for high priority pests that are thought to be exotic to the environment being tested
- false negatives lead to failure to detect pests that are present and are therefore more concerning for demonstrating freedom or early detection systems.

Validation of terrestrial and aquatic animal tests is a well understood process. There are many examples of well-validated assays in the published literature and of reporting standards, including for finfish, crustaceans and molluscs. Some of the methods and knowledge on animal disease diagnostics are also applicable to marine pest assays. However, there are many differences between animal disease and marine pest assays, posing significant challenges for the development and validation of new assays. Some of these differences and the resulting impacts on test validation approaches are summarised in Appendix A. Differentiating presence of agent (positive test result) from presence of disease (use of histopathology) is an important component of aquatic pest assays due to the nature of filter feeding.

5.2.1 Detection sensitivity

Detection sensitivity for marine pest assays is not well understood and is difficult to define and interpret. *Survey sensitivity* for visual pest surveys has been defined as 'the probability that a survey will detect the presence of a target species', and depends on the pest density (abundance), the number and volume of sampling units and the 'effectiveness' of the sampling units for detection of the pest if it is present in the sampling unit [Hayes et al. 2005].

For molecular assays for marine pests, the detection sensitivity of an assay is equivalent to the 'effectiveness' parameter described above. This can be thought of in two ways:

- The probability of a positive test result in a single sampling unit, given that the pest is present in the target environment and is affected by a range of factors, including pest abundance, distance from source, water currents, dilution, inhibition in the sample matrix, clustering and sampling (random) error.
- The probability of detecting a pest in a single sampling unit, given that the DNA is present in the sampling unit (for example plate/plankton tow) and is affected by a range of factors, including inhibition, clustering and sampling (random error).

Methods for estimating detection sensitivity are likely to vary depending on the interpretation being used. Evaluation at a site that is known to be infested is one solution, but samples need to be representative of the whole environment, not just from close to known pest locations to avoid overestimating sensitivity. Spatial relationships between pest locations and larval or environmental DNA detections are also poorly understood.

Spiking of samples is an alternative where other approaches are not possible but is less than ideal. Spiking of samples should be at dilutions close to the LOD, or at concentrations expected in the environment, into representative samples known to be otherwise free of the target.

Modelling approaches such as that described by Furlan et al. [2016] may also be a possibility in some circumstances, providing an estimate of overall sensitivity for a given water body for a given concentration of target ('survey sensitivity'), rather than unit-level estimates as would normally be the case.

5.2.2 Detection specificity

Detection specificity for marine pest assays can be interpreted as the proportion of negative reference samples (samples from a pest-free environment) that give a negative test result. Assays with a high analytical specificity should also have a high detection specificity. If this is not the case reasons for the lack of specificity should be investigated and the assay redesigned to improve specificity. Specificity estimates can be revised and improved as additional data becomes available from routine use of the assay in pest-free areas.

5.2.3 Reference samples

Evaluation of detection sensitivity and specificity often depends on the availability of reference samples of known status to allow a 'gold standard' evaluation. Such samples should be 'representative' of the environment and conditions in which the test may be used. For marine pests, this means that reference samples should be representative of the range of:

- environments in which the test may be used

- sample matrixes that may be submitted for testing
- seasonal conditions expected to occur
- likely pest abundance and proximity expected to occur.

Negative reference samples

Negative reference samples should be sourced from locations or times of year where/when the pest is confidently known to be absent from the local environment. Full details of sample type, sample matrix and volume, sampling location, timing and any other relevant details should be documented. Sampling known pest-free locations or times can raise issues of lack of representativeness, particularly for pests that are not exotic to Australia, because of the limitations of such locations and times compared to where and when the pest is likely to occur. Any such differences or limitations of reference samples should also be documented.

Positive reference samples

Positive reference samples should represent the spectrum of substrates, pest abundance and target analyte concentrations expected to occur in nature. True positive reference samples can be difficult to identify, particularly where the specific pest may be truly exotic from Australia or where there is no existing, well-characterised test available to confirm the status of the sample origin. In such cases, spiked samples may be used as an alternative to 'natural' samples.

Spiked samples

Spiked samples are commonly used for evaluating analytical performance of an assay but are less than ideal for evaluating detection performance. Natural environmental samples are preferred because of the greater similarity of composition to samples likely to be used in wider application of the assay. However, in the absence of positive reference samples, spiking of the target analyte (at concentrations expected in the environment, into representative environmental samples using biological material or extracted DNA) should be considered.

5.2.4 Test interpretation criteria

To enable interpretation of results for an assay which produces a qualitative result, the test result is usually reduced to two (positive and negative) or sometimes three (positive, negative and inconclusive/indeterminate) results. Test interpretation criteria for the assay should be defined and detection performance measures assessed based on the specified criteria.

For a quantitative test result this may require the selection of one (or two) thresholds or cut-points used to classify the results. Theoretically, setting a cut-point requires a compromise between assay sensitivity and specificity, depending on where the cut-point is set in relation to the assay's LOD. Selection of cut-points should consider the proposed purpose of the test and the relative importance of sensitivity and specificity to achieve that purpose.

Alternatively, for some assays, any normal curve amplification or other test response should be regarded as positive, and methods to re-test including using other assays, DNA clean-up and inhibition prevention should be attempted.

5.2.5 Sample sizes for estimating detection sensitivity and specificity

The sample size required for estimating detection sensitivity and specificity depends on the likely (assumed) value, the desired precision (confidence interval width) of the estimate and the evaluation method being used. The OIE *Aquatic Manual* [2016] provides guidance on sample size

required for evaluation against known infection status for animal disease assays, which would be similarly applicable for marine pest assays (see Appendix A).

As an example, 100 samples, either from affected environments or known spiked positive samples, and 300 samples from pest-free environments are generally considered the minimum sample size to provide an adequate level of confidence in the estimates of detection sensitivity and specificity, respectively. Assuming estimated sensitivity of 95 per cent (five false negatives), 95 per cent binomial confidence limits are 88.7 to 98.4 per cent. Similarly, assuming estimated specificity is 100 per cent (0 false positives), 95 per cent binomial confidence limits are 98.8 to 100.0 per cent.

Numbers required should be adjusted on a case-by-case basis, depending on the desired level of confidence and precision.

5.2.6 Estimation of detection sensitivity and specificity

Several methods are available for estimating detection sensitivity and specificity in terrestrial or aquatic animals. These methods may be applicable to varying degrees for evaluation of marine pest assays, as summarised below.

Based on results of testing reference samples

If well-characterised positive and negative reference samples are available, testing of these samples in the new assay will allow estimation of detection sensitivity and specificity. This is often problematic for animal tests and even more so for marine pest assays, due to the difficulty in reliably determining true status of the samples.

One exception to this is the use of negative reference samples from a pest-free location for estimation of detection specificity. As previously mentioned, such negative reference samples should be from a range of locations, environments, sample matrixes and seasons to maximise representativeness for the proposed use of the assay.

Alternatively, if there is an existing, well-characterised assay available, comparison of results with this assay is the preferred approach. This produces estimates of detection sensitivity and specificity, relative to the reference test.

Use of spiked samples of known status is an alternative to reference samples of known status; however, such samples will not be representative of the environment in which the assay will be used, limiting the applicability of the resulting estimates.

Evaluation at a known infested site

Evaluation of detection sensitivity of a new assay at a site (or sites) that is known to be infested with the target pest may be possible. A critical feature of such an evaluation would be to ensure spatial representativeness of samples across the site, rather than focusing the sampling in proximity to known pest occurrence. Some samples are likely to be more distant from pest locations and therefore have lower concentrations of the target analyte. Sampling only in proximity to known pest locations is likely to result in high concentrations of DNA and sensitivity could be overestimated, compared to sites with lower levels of infestation or further away from pest locations. Environmental characteristics, pest abundance, clustering and distribution should also be investigated and documented to assist in interpreting results and understanding test performance.

This approach will result in generally lower estimates of sensitivity than using reference or spiked samples, because of natural variation in the occurrence and concentration of the target analyte in

samples. However, such estimates are likely to be more indicative of the performance of the assay when used for its intended purpose.

Alternative modelling methods

Other modelling methods are available that may be applicable to the evaluation of marine pest assays. Specific methods that could be considered include:

- Simulation modelling of assay performance and detection as described by Furlan et al. [2016].
- Mixture modelling, or other latent class methods, where results from samples of unknown status from an infested environment are used to fit theoretical probability distributions for the underlying (and unidentified) true positive and true negative samples. These approaches are subject to a variety of important assumptions which must be met and generally require large numbers of samples to produce reliable estimates.

5.3 Reproducibility and augmented repeatability estimates

Reproducibility is the ability of an assay to provide consistent results when the same samples are tested using the identical assay (protocol, reagents and controls) in different laboratories and equipment. Reproducibility should be assessed with a panel of at least 20 samples, with identical aliquots, at a minimum of three laboratories. Data from these analyses can also be used to provide additional information on within-laboratory repeatability for participating laboratories.

With modern molecular assays, true reproducibility is often difficult to achieve due to the variety of platforms, reagents, chemistries and analysis used in different laboratories. To overcome this, harmonisation of methods and use of proficiency testing may be used as an alternative.

5.4 Assay implementation

The final test for a new assay is its deployment and use in the field, rather than in a research setting. The validation process described above provides initial insight into expected test performance, while additional data from field use provides valuable evidence that it is performing to expectations on an ongoing basis. Occasionally, initial field use will identify issues with an assay that require developers to re-evaluate assay performance and if necessary redesign and re-validate the assay to overcome identified issues. Ideally, assay implementation should include:

- development of an evaluation panel of standard samples to allow harmonisation between laboratories
- publication in a refereed journal.

5.4.1 Fitness for use

Before wider application of a new assay, it is important that the results of the validation process are used to assess the fitness of the assay for the intended purpose(s). This assessment should include not only the results of the validation process, but also consideration of other issues associated with successful implementation and adoption. These include acceptability by scientific communities, cost and acceptability to the client, and feasibility given available laboratory resources. An inability to meet operational requirements of an assay also may make it unfit for its intended use.

5.4.2 Interpretation of test results

Test results cannot be interpreted appropriately unless the detection performance of the test is well understood. Accordingly, developers should publish or otherwise inform potential users of the assay

of performance specification, particularly estimates of detection sensitivity and specificity. Once armed with this information, users can interpret assay results to assist in decision-making, for example on whether further follow-up of positive results is required and what form that follow-up should take.

5.4.3 Deployment to other laboratories

An important part of implementation is the deployment of the assay to other laboratories. This encourages wider use and allows improved monitoring of performance through proficiency testing and ongoing assessment of assay reproducibility. Preparation and distribution of a sufficient volume of well-characterised and consistent reference standards, to support wider use and ensure consistency of results between laboratories, is a critical consideration for wider deployment.

5.5 Monitoring assay performance after initial validation

Monitoring and maintenance of an assay should include, at a minimum, development of a proficiency testing scheme for participating laboratories. Further, participation in a quality control programme, where a programme exists, is mandatory under National Analytical and Testing Authority standards.

5.5.1 Monitoring the assay

Once a new assay has been validated, it is important to maintain consistent validated performance characteristics. A quality assurance programme should assess the daily performance of the assay, primarily through monitoring of precision and accuracy estimates for internal controls, as well as monitoring of outlier tendencies.

5.5.2 Considerations for changes in the assay

It may be necessary to periodically modify an existing assay, for example to meet a new intended purpose, to take advantage of technological advances or to improve efficiency or cost-effectiveness. When this occurs, revalidation may be appropriate depending on the type and magnitude of changes. Revalidation should be considered for changes to the intended purpose(s) or the target environment, or to the target analyte for the assay.

Technical modifications and comparability assessments

Minor technical modifications to a validated assay, such as changes in instrumentation, extraction protocols, or conversion of an assay to a semi-automated or fully automated system using robotics, will typically not necessitate full revalidation of the assay. Instead, a comparison of the current and proposed methods should be undertaken to determine whether the proposed modifications will affect the documented performance characteristics of the assay.

Biological modifications and comparability assessments

Changes to the biological components of an assay are more difficult to assess. These may include changes to the sample matrix or the reagents used. Where biologicals are changed, a final decision on whether a full revalidation is required may be based on a re-assessment of the analytical performance, and the comparison of new and old biologicals on the performance characteristics of the assay.

Replacement of depleted reagents

When a reagent such as a control sample or working standard is nearing depletion, it is essential to prepare and repeatedly test a replacement before such a control is depleted. The prospective control sample should be included in multiple runs of the assay in parallel with the original control to

establish their proportional relationship. It is important to change only one reagent at a time to avoid the compound problem of evaluating more than one variable.

5.5.3 Enhancing confidence in validation criteria

As opportunity arises, it is desirable to increase the number and range of reference samples, to allow recalculation of updated and improved estimates of detection performance of the assay. These samples should be collected and managed under conditions as similar as possible to the original samples to ensure comparability of results.

6 Reporting and documentation

To fully validate a new assay, developers and new users should address all the validation criteria:

- definition of the intended purpose(s)
- optimisation
- standardisation
- repeatability
- analytical sensitivity
- analytical specificity
- test interpretation criteria
- detection sensitivity
- detection specificity
- reproducibility
- fitness of use
- suitability of multiplex tests (to detect related organisms).

Documentation should include (at a minimum) full description of:

- target pest(s) and assay target(s)
- assay methods (and variations used in the validation process)
- platforms and equipment used, including collaborating laboratories
- sample types and sampling details for reference samples
- other samples used for validation, including sample matrix, volume, composition, source location and time, spiking concentrations and methods (if appropriate) and any other relevant sample details
- methods used
- results for each of the validation criteria.

The assay should not be considered validated for any sample type, method variation, equipment or other change to the sampling and testing process unless the variation has either been included in the validation process or equivalency has been demonstrated and documented.

Summary

An appropriately optimised and validated assay will provide consistent results that can be interpreted by the user, resulting in improved understanding of pest presence/absence or abundance and improved decision-making in relation to implementation of control or eradication measures, if required. Without proper validation, tests are likely to provide inconsistent or incorrect results, leading to poor decisions and loss of faith in the testing system.

Appendix A: Sample size estimates

Table A1 Estimated sample size required for estimating detection sensitivity (DSe) or detection specificity (DSp), depending on expected value of DSe or DSp and desired precision and confidence

Estimated DSe or DSp	Precision ± 2%			Precision ± 5%		
	90% confidence	95% confidence	99% confidence	90% confidence	95% confidence	99% confidence
90%	610	864	1493	98	138	239
92%	466	707	1221	75	113	195
94%	382	542	935	61	87	150
95%	372	456	788	60	73	126
96%	260	369	637	42	59	102
97%	197	279	483	32	45	77
98%	133	188	325	21	30	52
99%	67	95	164	11	15	26

DSe = detection sensitivity. DSp = detection specificity.

Appendix B: Differences between animal disease and marine pest testing

Validation of marine pest assays presents significant challenges compared to assays for terrestrial or aquatic animals. These challenges relate to some fundamental differences arising from the fact that for animal diseases, tissue or fluid samples can usually be obtained directly from individual animals for testing, whereas marine pest diagnostics often rely on sampling from the marine environment and, less commonly, directly from the organism. Some of these important differences are summarised in Table A1. Resulting issues relating to current approaches to estimation of diagnostic sensitivity and specificity of animal and marine pest diagnostics are summarised in Table A2.

Table B1 Summary of comparisons between animal and marine pest assays

Issue	Terrestrial/aquatic animal disease assays	Marine pest assays
Sampling unit	Usually a visual animal	Can be highly variable and can be any of a variety of environmental samples, such as a volume of water, sediment cores, settlement plates, plankton tows, etc.
Population of interest	An aggregation of animals in a herd, farm, pond, tank, water body or at a higher level of regional or national populations	Usually a harbour, port or other water body. Distribution of the assay target is unlikely to be uniform and is affected by dilution, distance from source, currents and other factors
Sampling strategies for application of the test	Can be well developed and common across species and diseases	Less well developed, highly variable depending on assay matrix and target organism
Sample matrix	Usually blood, tissue or other body fluids/excreta. Often homogeneous and predictable in physical/chemical properties	Highly variable in nature depending on the type of pest/target and in time and space. Often a very complex biological substrate, such as ocean sediment, plankton tows and settlement plate scrapings
PCR Inhibition	Depends on sample matrix and usually manageable	Commonly associated with the complex nature of the sample matrix
Number of assay targets	Usually a single target of interest, associated with a particular disease. Multiplex are usually for a limited number of targets	Often multiple targets due to the variety of pests of concern. Also, often need to distinguish between multiple closely related species that may or may not be of interest. The taxonomic framework around some of these species is poor
Concentration of assay target	Can be high, depending on infectious dose and on nature of the specific disease and test	Often high, making false negatives likely, depending on concentration and distribution of the target in the water body
Sample volume	Usually small, can be measured in fractions of a gram or ml	Often very large, particularly where concentration of the target DNA is low
Availability of comparison/reference tests	Multiple alternative and comparison tests may be available	Rarely are good alternative tests available
Species diversity	Usually single species	May be single or multiple species, may be related species but also a need for broader coverage of multiple pests in a single assay or series of assays. Ideally able to distinguish between closely related species

Representativeness of evaluation samples	Evaluation samples need to be representative of the population in which test is to be used and of the spectrum of disease. This is achieved with variable success but is at least achievable for many tests	Representativeness requires samples from different times of year, sample matrixes and geographic locations, with associated logistic and cost issues. Replication of assay performance can be difficult due to different environments and complex and varying sample composition

PCR Polymerase chain reaction.

Table B2 Summary of issues affecting validation methods for animal disease and marine pest assays

Issue	Terrestrial/aquatic animal diseases	Marine pests
Gold standard methods	Gold standard methods applicable and straightforward when an appropriate reference test is available for comparison	Gold standard methods unlikely to be applicable due to lack of alternative tests as a gold standard. Novel molecular tests often substantially more sensitive than existing alternative techniques (visual inspection, trapping, etc.). Gold standard approaches may be of use in establishing specificity estimates if known pest-free environments or multiple assays are available
Latent class methods	Latent class methods are increasingly popular for animal diagnostics	Unlikely to be a realistic option for most marine pest assays due to methodological requirements (multiple tests, multiple populations, constant test performance, etc.)
Spiked samples	Usually limited to preliminary evaluation and estimating analytical sensitivity and specificity	Commonly used for estimating analytical sensitivity and specificity, may also be applicable for estimating diagnostic sensitivity in samples around the limit of detection, or at concentrations expected in the environment. Sample matrix for spiked samples should be free of the target of interest prior to spiking
Novel modelling methods	May be applicable but not commonly used	May be opportunities for some pests/assays, for example as described by Furlan. See Furlan et al, [2016] for more information

Glossary

Term	Definition
Accuracy	The closeness of a test value to the expected (true) value (mean or median) for a reference standard reagent of known concentration
Analytical sensitivity	Synonymous with 'limit of detection'; smallest detectable amount of analyte that can be measured with a defined certainty; analyte may include antibodies, antigens, nucleic acids or live organisms
Analytical specificity	Degree to which the assay distinguishes between the target analyte and other components in the sample matrix; the higher the analytical specificity, the lower the level of false positives
Coefficient of variation	The ratio of the standard deviation of a set of sample measurements to their mean value
Detection sensitivity	Proportion of either: 1) known 'positive' reference samples, or 2) samples from a known infested environment, that test positive in the assay. Comparable to diagnostic sensitivity in animal disease assays
Diagnostic specificity	Proportion of either: 1) known 'negative' reference samples, or 2) samples from a known uninfested environment, that test negative in the assay. Comparable to diagnostic specificity in animal disease assays
Optimisation	The process by which the most important physical, chemical and biological parameters of an assay are evaluated and adjusted to ensure that the performance characteristics of the assay are best suited to the intended application
Precision	The degree of dispersion (variance, standard deviation or coefficient of variation) within a series of measurements of the same sample tested under specified conditions
Repeatability	Level of agreement between replicates of a sample both within and between runs of the same test method in a given laboratory
Reproducibility	Ability of a test method to provide consistent results when applied to aliquots of the same sample tested by the same method in different laboratories
Robustness	An assay's capacity to remain unaffected by minor variations in test situations that may occur over the course of testing in a single laboratory

References

Furlan EM, Gleeson D, Hardy CM & Duncan RP 2016, 'A framework for estimating the sensitivity of eDNA surveys', *Molecular Ecology Resources*, vol. 16, no. 3, pp. 641–54.

Hayes KR, Cannon R, Neil K & Inglis G 2005, 'Sensitivity and cost considerations for the detection and eradication of marine pests in ports', *Marine Pollution Bulletin*; vol. 50, no. 8, pp. 823–34.

OIE 2016. Manual of Diagnostic Tests for Aquatic Animals. World Organisation for Animal Health (OIE), Paris, France, 2016.

OIE 2016a. Manual of Diagnostic Tests and Vaccines for Terrestrial Animals. World Organisation for Animal Health (OIE), Paris, France, 2016.