A validation scale to determine the readiness of environmental DNA assays for routine species monitoring

Bettina Thalinger[1,2]*, Kristy Deiner[3,4], Lynsey R. Harper[5,6], Helen C. Rees[7], Rosetta C. Blackman[8,9], Daniela Sint[2,10], Michael Traugott[2,10], Caren S. Goldberg[11] & Kat Bruce[12]

[1] Centre for Biodiversity Genomics, University of Guelph, 50 Stone Road E, N1G 2W1, Guelph, Ontario, Canada

[2] Department of Zoology, University of Innsbruck, Technikerstr. 25, 6020, Innsbruck, Austria

[3] Natural History Museum, Cromwell Road, London, SW7 5BD, UK

[4] Institute of Biogeochemistry and Pollutant Dynamics, ETH Zurich, Universitätstr. 16, 8092 Zurich, Switzerland

[5] Illinois Natural History Survey, Prairie Research Institute, University of Illinois at Urbana-Champaign, Champaign, IL, USA

[6] School of Natural Sciences and Psychology, Liverpool John Moores University, Liverpool, L3 3AF, UK

[7] RSK ADAS Ltd, School of Veterinary Medicine and Science, The University of Nottingham, Sutton Bonington Campus, Loughborough, LE12 5RD, UK

[8] Eawag: Swiss Federal Institute of Aquatic Science and Technology, Department of Aquatic Ecology, Überlandstr. 133, 8600 Dübendorf, Switzerland

[9] Department of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstr. 190, CH-8057 Zurich, Switzerland

[10] Sinsoma GmbH, Lannes 6, 6176 Völs, Austria

[11] School of the Environment, Washington State University, Pullman, WA 99164, USA

[12] Nature Metrics Ltd, CABI Site, Bakeham Lane, Egham, TW20 9TY, UK

***Corresponding author:**

Bettina Thalinger, bettina.thalinger@gmail.com

Centre for Biodiversity Genomics, University of Guelph, 50 Stone Road E, N1G 2W1, Guelph, Ontario, Canada

Running headline: eDNA assay validation scale

Abstract

The use of environmental DNA (eDNA) analysis for species monitoring requires rigorous validation - from field sampling to interpretation of PCR-based results - for meaningful application and interpretation. Assays targeting eDNA released by individual species are typically validated with no predefined criteria to answer specific research questions in one ecosystem. Their general applicability, uncertainties and limitations often remain undetermined. The absence of clear guidelines prevents targeted eDNA assays from being incorporated into species monitoring and policy, thus their establishment will be key for the future implementation of eDNA-based surveys. We describe the measures and tests necessary for successful validation of targeted eDNA assays and the associated pitfalls to form the basis of guidelines. A list of 122 variables was compiled, consolidated into 14 thematic blocks, such as "*in silico* analysis", and arranged on a 5-level validation scale from "incomplete" to "operational". Additionally, minimum validation criteria were defined for each level. These variables were evaluated for 546 published single-species assays. The resulting dataset was used to provide an overview of current validation practices and test the applicability of the validation scale for future assay rating. The majority (30%) of investigated assays were classified as Level 1 (incomplete), and 15% did not achieve this first level. These assays were characterised by minimal *in silico* and *in vitro* testing, but their share in annually published eDNA assays has declined since 2014. The total number of reported variables ranged from 20% to 76% and deviated both between and within levels. The meta-analysis demonstrates the suitability of the 5-level validation scale for assessing targeted eDNA assays. It is a user-friendly tool to evaluate previously published assays for future research and routine monitoring, while also enabling appropriate interpretation of results. Finally, it provides guidance on validation and reporting standards for newly developed assays.

Key words: digital PCR, eDNA, endpoint PCR, https://edna-validation.com/, quantitative PCR, species-specific, species detection

1. Introduction

Determining the occurrence of species is essential for ecology and requires sensitive and accurate detection methods. Within the last decade, species detection from environmental DNA (eDNA; i.e. detection of trace DNA released by organisms into their environment) has shown great potential for routine species surveys (Rees, Maddison, Middleditch, Patmore, & Gough, 2014; Goldberg et al., 2016; Deiner et al., 2017). The interest in molecular species detection has fuelled the development of over 500 assays, reviewed herein, that utilise PCR of DNA or RNA extracted from environmental samples. Generally, "targeted" eDNA assays must be specific to the species of interest and possess high sensitivity to allow detection at low densities, low DNA concentrations, and across spatiotemporal scales (Goldberg et al., 2016; MacDonald & Sarre, 2017).

A targeted eDNA assay encompasses the entire workflow used to detect a species' DNA from an environmental sample, inclusive of field sampling to the interpretation of PCR-based results, and not just the primers and probes. Thus, adherence to workflows will determine the success or failure of an eDNA assay because methodological choices influence performance and sensitivity (e.g. Doi, Takahara, et al., 2015). An assay is often validated within a specific system to answer a set question about the target species. Its applications beyond initial development are hampered by the poor understanding of remaining uncertainties, such as the potential for false positives resulting from non-target amplification or contamination, or false negatives resulting from low sensitivity, sample degradation, low DNA yield protocols or inhibition (Goldberg et al., 2016; Lacoursière-Roussel & Deiner, 2019). In a management context, false positives and false negatives may lead to misuse of resources (e.g. funds and personnel) for issues such as rare species protection and invasive species control. Both scenarios foster inaccurate interpretation of results, fuelling arguments against routine use of eDNA detection for species monitoring (Jerde, 2019).

Aside from a few well-validated eDNA assays that have been developed and incorporated into routine monitoring, the application of published assays is a minefield for end-users to navigate. We illustrate this through two examples. The assay for great crested newt (*Triturus cristatus* (Laurenti, 1768), a legally protected species in the UK [Natural England, 2015]) was one of the first eDNA assays validated in both laboratory and field trials against conventional tools, demonstrating its potential for routine monitoring (Thomsen et al., 2012; Rees, Bishop, et al., 2014). After successful validation, a national eDNA-based citizen science monitoring scheme was tested and showed that large-scale eDNA sampling can enable distribution modelling (Biggs et al., 2015). These initial studies paved the way for eDNA-based *T. cristatus* detection to inform new policies aimed at providing landscape-level species protection (L. R. Harper, Buxton, et al., 2019). Studies have since investigated optimal methods of eDNA capture, relative abundance and detection probability estimation, and the influence of seasonality as well as biotic and abiotic factors on *T. cristatus* eDNA detection and quantification (Buxton, Groombridge, & Griffiths, 2017; Buxton, Groombridge, Zakaria, & Griffiths, 2017; Buxton, Groombridge, & Griffiths, 2018a, 2018b). Due to these combined efforts, the assay has undergone exemplary validation and is operational for management.

Conversely, no assays have been successfully applied to routine monitoring of invasive American crayfish. For example, there is a lack of consensus on a single assay for the signal crayfish (*Pacifastacus leniusculus* (Dana, 1852)). Larson et al. (2017) developed and tested an assay against conventional trapping, but five other assays have also been proposed with differing degrees of validation and divergent *in silico* and *in vitro* approaches (Agersnap et al., 2017; Dunn, Priestley, Herraiz, Arnold, & Savolainen, 2017; K. J. Harper, Anucha, Turnbull, Bean, & Leaver, 2018; Mauvisseau et al., 2018; Robinson, Uren Webster, Cable, James, & Consuegra, 2018). The *P. leniusculus* assays were developed using a variety of strategies for eDNA sampling, capture, extraction, qPCR of different genetic markers, and applied across genetically diverse populations within the species range. Due to this substantial methodological

variability, direct comparisons between results obtained from these assays are impossible. Therefore, the *P. leniusculus* assays represent a minefield for end-users, despite the need for accurate and sensitive tools to enable actionable species management of this invasive species in Europe.

These case studies exemplify how consensus and dissent in assay validation can influence the implementation of eDNA analysis for species monitoring. Developing guidelines to determine the suitability of eDNA assays for end-users will therefore ensure that ecological insights or management decisions are based on robust molecular analyses with quantifiable uncertainties and clear inference limits (Goldberg et al., 2016; MacDonald & Sarre, 2017; Nicholson et al., 2020). Here, we describe the general validation process for targeted PCR-based methods and examine the extent of assay validation and reporting in the eDNA literature. We present an eDNA assay validation scale, which establishes criteria to enable the classification of assays based on their accuracy and sensitivity for single-species detection. To demonstrate the utility of the scale, we performed a meta-analysis of targeted eDNA assays published in 327 papers as of 11th April 2019 (546 assays). By placing an eDNA assay on the validation scale, end-users can determine the recommended scenarios for application and improve assay performance with further validation.

2. Criteria and principles of validation

2.1. General requirements for an eDNA laboratory

All laboratory activities are subject to error. In order to have confidence in results, quality standards and good practices are required in diagnostic laboratory environments (e.g. World Health Organization, 2011; Halling, Schrijver, & Persons, 2012). Although few eDNA-processing

laboratories will employ 'ancient DNA' practices (e.g. hazmat suit and positive air pressure with HEPA filtered inflow), all laboratories conducting eDNA analysis should utilise a unidirectional workflow where pre-PCR steps are performed in separate laboratories dedicated to low DNA quality and quantity (Goldberg et al., 2016). Completely standardised laboratory environments are rare and the use of proficiency tests (as conducted by UK laboratories participating in *T. cristatus* monitoring) can help end-users understand the quality of results obtained among different laboratories. Even results obtained from an extensively validated assay can be questionable when they are not produced within a suitable laboratory environment (Goldberg et al., 2016).

2.2. Reporting standards for in silico*,* in vitro *and* in situ *validation of assays*

Targeted eDNA assay validation is a multi-step process. It can be divided into *in silico* validation (i.e. computer-based tests for primer specificity), *in vitro* validation (i.e. laboratory tests with reference tissue samples) and *in situ* validation (i.e. field tests with eDNA samples) (see Goldberg et al., 2016; MacDonald & Sarre, 2017). Understanding the utility of an assay requires both knowledge of the context in which it has been designed, and a broader understanding of how it was developed. Here, we give a brief overview of what key steps comprise assay validation, with a focus on troubleshooting steps that may be necessary when applying previously published assays to new locations or with modified laboratory practices (Table 1).

The first step is *in silico* assay validation, the goal of which is to determine assay specificity based on known sequence diversity. Sequence diversity has three categories, including sequences from: (i) closely related and co-occurring species, (ii) closely related but geographically distinct species, and (iii) distantly related but co-occurring species (i.e. sequences that could co-amplify and produce false positive results for a target species). By checking primer specificity from available sequences, the geographic area of applicability for an

assay can be maximised through identifying and removing potential issues of co-amplification. Typically public or custom databases are used for performing *in silico* amplification (e.g. ecoPCR [Boyer et al., 2016]; primerBlast [Ye et al., 2012]; PrimerTREE [Cannon et al., 2016]; PrimerMiner [Elbrecht & Leese, 2017]). While reference sequence libraries are often far from complete and many of the factors influencing successful PCR amplification cannot be simulated, *in silico* testing provides a first impression of primer performance and should be conducted.

The essential components of *in vitro* assay validation are optimisation, specificity, and sensitivity. Tests with varying PCR chemistry, reaction volume, primer/probe concentration, cycling conditions and technical replication will ensure optimal, standardised and error-free target DNA amplification (Bustin et al., 2009; Wilcox, Carim, McKelvey, Young, & Schwartz, 2015; Goldberg et al., 2016). The assay must then be tested against closely related and co-occurring non-target taxa to ensure specificity (Goldberg et al., 2016). Ideally, tissue-derived DNA samples from multiple individuals spanning a defined geographic area are tested to ensure the assay is robust to genetic variants of target and non-target species. Amplicons should be Sanger sequenced to confirm species identity (Goldberg et al., 2016), but short fragments (<100 bp) have limited sequence length available for species determination (Meusnier et al., 2008). Next, the Limit of Detection (LOD) must be determined to assess assay sensitivity, and the Limit of Quantification (LOQ) determined, if the measurement of eDNA quantity is desired. Generally, these values are obtained using a dilution series of quantified DNA amplicons or synthesized gene fragments (e.g. gBlock, IDT) based on public or *de novo* reference sequences. The LOD and LOQ have various definitions in the eDNA literature, but were recently standardised by Klymus et al. (2019), where LOD is the lowest standard concentration at which 95% of technical replicates amplify and LOQ is the lowest standard concentration for which the coefficient of variation (CV) value is <35%. Unfortunately, the existence of past definitions requires the final LOD and LOQ to be reported as well as the definition used. We note that these metrics apply directly to the assay as developed and assume no interference during PCR

from the rest of the species' genome (i.e. if a gBlock is used), other genetic material, or inhibitory compounds.

Finally, the assay must be validated *in situ* by surveying sites with and without the target species (Goldberg et al., 2016). It must be tested against conventional tools for presence/absence detection and tests for estimation of relative abundance/biomass are advisable. Assays are deemed successful if eDNA and conventional detections concur at occupied sites and no eDNA detections are observed at definitively unoccupied sites. Sanger sequencing of eDNA amplicons can provide additional evidence but cannot distinguish sample contamination from true detections (Goldberg et al., 2016). Besides screening for the target species, negative eDNA samples (or all eDNA samples if quantification is necessary) should be tested for inhibition. This requires an Internal Positive Control (IPC) assay for synthetic DNA (e.g. ThermoFisher) or an assay for non-target species using exogenous or endogenous DNA (Goldberg et al., 2016; Doi et al., 2017; Furlan & Gleeson, 2017).

Advanced *in situ* validation may investigate the influence of biotic (e.g. abundance, biomass, life stages, microbial activity) and abiotic (e.g. temperature, pH, ultraviolet light, salinity) factors influencing eDNA origin, state, fate, and transport (Barnes & Turner, 2016; Lacoursière-Roussel & Deiner, 2019). Assays that account for spatial (e.g. shoreline versus offshore) and temporal (e.g. summer versus winter) variation in eDNA distribution and abundance due to the ecology of a species can be implemented with greater confidence (de Souza, Godwin, Renshaw, & Larson, 2016; Lawson Handley et al., 2019). Occupancy modelling using eDNA data is desirable as it accounts for detection probability while estimating site occupancy, even if all field samples from a site return negative. Hierarchical models that incorporate eDNA occupancy and detection probabilities at site, sample, and technical replicate levels are most accurate and can be implemented in software such as R (e.g. package "eDNAoccupancy" [Dorazio & Erickson, 2018]) or PRESENCE (MacKenzie et al., 2002). However, model assumptions regarding false positives should be carefully considered.

3. Types and trade-offs of targeted eDNA detection methods

Amid the processing chain (i.e. sampling to data analysis) for a targeted eDNA assay, PCR warrants extra consideration as the technological spectrum and potential for variation is enormous. Previous publications have typically defined an assay as the primers (and probe) required for DNA amplification and associated visualisation (e.g. agarose gel electrophoresis, qPCR software). Hence, differences between the multiple detection instruments and chemistry used in combination with species-specific primer (and probe) sets are at the core of targeted eDNA assays. Table 2 provides an overview of amplification types and their associated trade-offs.

Many assays - especially those published in earlier years - use endpoint PCR. However, most assays to date employ real-time quantitative (q)PCR allowing for greater sensitivity and quantitative data. More recent publications have used digital (d)PCR for absolute quantification. Alternatively, LAMP and CRISPR were shown to be suitable for eDNA applications, decreasing the requirements of in-field testing equipment (M. R. Williams et al., 2017; M. A. Williams et al., 2019). A few publications use alternative methods such as PCR combined with restriction fragment length polymorphism (RFLP). All amplification types (Table 2) enable distribution and occupancy modelling, provided enough biological and technical replication is employed (Hunter et al., 2015; Goldberg, Strickler, & Fremier, 2018; Wilcox et al., 2018). However, endpoint PCR in combination with agarose gels is the only type where a detection limit cannot be set objectively (low sensitivity) and which does not provide estimates of DNA copy number (Doi, Uchii, et al., 2015; Yamanaka & Minamoto, 2016; Hunter et al., 2017; Thalinger, Wolf, Traugott, & Wanzenböck, 2019). Depending on the management context and study size, the optimal

242 detection instrument can vary, albeit technological advances will slowly shift the focus away
243 from endpoint PCR.
244
245
246 **4. Evaluating the current status of assay validation**
247
248 To understand current validation practices for targeted eDNA assays, we generated an
249 extensive list of variables deemed important for assay validation by 35 experts in the field of
250 targeted eDNA detection whom convened at a DNAqua-Net EU COST Action (Leese et al.,
251 2016) workshop held on 26-27 March 2018 at the University of Innsbruck, Austria. The variables
252 in this list consist of 18 categorical variables (e.g., species identity, target gene, sample type)
253 and 104 binomial variables directly associated with the eDNA processing chain, from primer
254 design to interpretation of field study results (SI1).
255 A comprehensive literature database for targeted eDNA assays was built in three steps.
256 First, we included all papers listed on the 'eDNA assays' web page
257 (https://labs.wsu.edu/edna/edna-assays/) as of 10 April 2019. Second, we conducted a Web of
258 Science literature search on 11 April 2019, including the search terms "environmental DNA" and
259 "eDNA" but excluding terms associated with microbial organisms and metabarcoding (SI2A).
260 Third, the resulting 660 Web of Science entries were manually checked for suitability (i.e.
261 macrobial target organism, targeted eDNA detection intended) leading to a combined database
262 of 327 papers. For each of the assays contained in these papers, the 122 variables were
263 recorded in a checklist by one of the authors. Before data entry, all authors validated the same
264 four papers (Deiner & Altermatt, 2014; Rees, Bishop, et al., 2014; Thalinger et al., 2016; L. R.
265 Harper, Griffiths, et al., 2019) to ensure recorder standardisation. Validation efforts were
266 classified as 1 for "tests done, or parameter reported", 0 for "variable not reported, or no testing
267 done", and NA in cases where the respective variable did not match the assay type (e.g. filter

type when samples were precipitated). When an assay was used in multiple papers, all validation efforts were summarised in one entry and the literature database was extended with the papers reporting primer sequences or other methodological aspects. As the type of amplification is important for assay validation, primer pairs used on multiple detection platforms were given separate database entries per amplification type. However, because most assays developed were presented in one publication, we did not account for slight variations in other aspects of the workflow (e.g. different extraction method, different filter type). After recording the values for each eDNA assay using the validation checklist (SI1), each author scored the assay intuitively based on a preliminary version of the validation scale (see section 5). The resulting database of 122 variables for each assay was the basis for all further analyses using R (R Core Team, 2019) and associated packages (SI2B).

Altogether, 546 assays from 327 papers were assessed. Of these assays, 227 were designed to detect fish species and 74 were designed to detect amphibian species; hence, it is unsurprising that ~80% of assays utilised water sampling. Fourteen percent of the assays were tested on tissue only and few assays were optimised for other sample types such as aerosol, sediment, snow or soil. More than 80% of assays were reported in only one paper, and most were designed for qPCR (~60%) or endpoint PCR (~35%) platforms. The cytochrome c oxidase subunit I (*COI*) gene was the most popular (>40%) genetic marker, followed by the cytochrome b (*cytb*) gene (~23%) (SI3).

5. The 5-level assay validation scheme

To enable standardised assay validation and reporting in the future, we assigned the assay-specific variables to 14 thematic blocks such as "*in silico* analysis", "PCR", or "extensive field testing of environmental samples" (Table 3, Fig. 1). These blocks were placed on a five-level

scale enabling the categorisation of assays from Level 1 ("incomplete") to Level 5 ("operational"), and the interpretation of associated field study results (Fig. 1). Each of these blocks contained variables associated with either rudimentary or substantial validation and reporting. For instance, the thematic block "concentration of eDNA from environmental sample" contains the variable "volume/weight of environmental sample", which was reported for almost all assays, but also contains "pressure used for filtration", which was rarely reported and/or measured. Therefore, a minimum criterion was introduced for each variable block functioning as proof of validation. For example, "detection from an environmental sample" was used as evidence that some validation had been undertaken in the block "detection obtained from environmental samples" (Table 3). To reach a level on the validation scale, an assay must fulfil all minimum reporting criteria for that level. For each assay, we calculated a total scoring percentage and a block scoring percentage for the 14 blocks by dividing the number of variables tested or reported by the complete set of variables possible. The total possible for each percentage included only variables relevant to the applied methods (e.g., for assays using filtration, precipitation variables were omitted; see Box 1 for example assays).

Of the 546 assays analysed, the majority (30%) were classed as Level 1. Of the remainder, 15% ($N = 83$) did not fulfil the minimum criteria necessary to reach Level 1, and no assay reached Level 5 (Fig. 1 and Fig. 2A). Newer assays published after 2016 were more likely to reach Level 4, and the percentage of assays failing to reach Level 1 gradually declined since 2014 (Fig. 2B). Generally, the total scoring percentage for all variables in a level increased from Level 0 to Level 4, but its variation was non-uniform with most outliers observed at Level 2 (Fig. 2C). Assays not reaching Level 1 exhibited scoring percentages between ~20% and ~55%, clearly showing the difference between incomplete and partially validated assays, which did not achieve higher levels due to one or several missing validation step(s). Generally, variables associated with lower levels on the validation scale were more likely to be reported or tested.

Nevertheless, some variables (e.g. “haplotypes of target tissue” or “pressure used for filtration”), were addressed by fewer than 10% of assays (Fig. 2D).

The rigour of the minimum criteria was evaluated by tallying the not achieved or reported cases. Specifically, 62 assays did not reach Level 1 because the targeted species sequence(s) used in primer design were not reported, which is a prerequisite for this level. Eight assays did not report the primer sequence. The lack of target detection from an environmental sample and omission of filter type or precipitation chemicals were the most restrictive criteria and were not fulfilled by 80 and 43 assays respectively, most of which were exclusively used for tissue tests (Fig. 2E). For assays ranked at Levels 2 and 3, there was agreement between the intuitive assay rating provided by the recorder and that assigned by the objective criteria. For assays placed at Level 1 following objective criteria, authors tended to be more liberal and rated them one or two levels higher (SI4). Finally, a classification tree analysis (De’Ath & Fabricius, 2000) was carried out to identify common characteristics of assays placed at each level of the validation scale (SI5). Most assays failing to reach Level 1 showed distinctly low levels of *in silico* validation. This was also true for most Level 1 assays (N = 103), albeit these showed higher levels of target tissue validation (Fig. 3). On the other end of the spectrum, Level 4 assays were characterised by substantial testing or reporting for *in vitro* testing, field sample processing, LOD determination, PCR, and field testing.

6. Conclusions and outlook

The validation scale and reporting standards developed here are an inclusive set of guidelines for targeted eDNA assays. They are directed at both the scientific community and end-users who want to apply a previously published assay or develop and publish a new assay. One needs to acknowledge that a strict standardisation of eDNA assays will not be possible due to

345 applications for manifold taxa in diverse ecosystems combined with technological advances. By
346 checking which of the 122 validation variables have been addressed, it is possible to identify
347 both available and missing information needed to successfully develop or reuse an assay. As a
348 general recommendation, authors should report as much information as possible on the
349 conducted validation steps, either in the main text or in the supplementary material. Additionally,
350 developers should consult existing guidelines for best practices along the validation workflow
351 prior to assay design and fieldwork (Bustin et al., 2009; Goldberg et al., 2016; MacDonald &
352 Sarre, 2017; Klymus et al., 2019). To make the validation process accessible, we provide a
353 checklist in SI1. Furthermore, a website (https://edna-validation.com/) was created to
354 summarise the cornerstones of the validation process and the validation scale. It will be possible
355 to enter all 122 variables and the minimum criteria to rank assays and calculate their scoring
356 percentages on the scale. The website will also serve as a living document when improvements
357 in technology and/or our understanding of eDNA in the environment advance.

358 The 5-level validation scale designed here provides an overview of the capabilities and
359 uncertainties of targeted eDNA assays. However, the binary data entry system cannot replace a
360 close check of previous publications as it does not always allow a qualitative assessment.
361 Details for validation variables are often spread across different sections in a publication or
362 ambiguously displayed. Thus, the checklist can be used as standard reporting guidelines for
363 targeted eDNA assays. It should be emphasised that for specific research questions and
364 associated publications, minimal validation efforts may be sufficient. Nevertheless, thorough
365 validation is needed to reduce uncertainties and overcome the limitations associated with
366 eDNA-based species monitoring. Furthermore, it is important that practitioners consider how an
367 assay can be modified (e.g. using different PCR reagents) and whether this changes its
368 validation level.

369 The successful application of targeted eDNA assays for routine species detection and
370 monitoring largely depends on the scientific community and the industry providing eDNA

services. Laboratories participating in ring tests such as that proposed for metabarcoding (Blackman et al., 2019) can facilitate consensus on analysis standards. For now, assay developers must respond to queries and help troubleshoot reproducibility issues. Such engagement will facilitate the application of targeted eDNA assays by other users and outside their original geographic scope or academic context. Finally, it is necessary for both the scientific community and the commercial laboratories to communicate realistic applications and limitations to end-users, as often an assay is not bad *per se*, but simply unsuitable.

Acknowledgements

This article is based upon work from COST Action DNAqua-Net (CA15219), supported by the COST (European Cooperation in Science and Technology) programme. We thank the workshop participants who convened on 26-27 March 2018 at the University of Innsbruck, Austria, where initial brainstorming for this manuscript occurred. We acknowledge additional guidance and input from Micaela Hellström, and are grateful to Cathryn Abbott for the improved design of Fig.1. We thank Nature Metrics Ltd. for support in hosting the independent website formed to operationalise the validation scale for future use.

Author's contributions

KB conceived the original idea which was developed by all co-authors. BT compiled the literature database and BT, KD, LRH, HCR, RCB, DS, MT, and KB collected the data. BT analysed the data and led the writing of the manuscript for which first drafts of sections were

396 provided by LH, RCB and KD. All authors contributed critically to the drafts and gave final
397 approval for publication.
398
399
400 **Data availability**
401
402 The literature database, the validation checklist compiled from it, the R code, and seven
403 datasets derived from the checklist and used for analyses are available on Figshare at
404 https://doi.org/10.6084/m9.figshare.12184860.v1 The literature database, the validation scale,
405 and the full checklist are also available at https://edna-validation.com/.
406
407
408 **Conflict of interest**
409
410 KB is the co-founder and CEO of Nature Metrics Ltd. a for profit company dedicated to the
411 analysis of environmental DNA. HCR manages environmental DNA services for RSK ADAS Ltd.
412 a for profit environmental consultancy. DS and MT are co-founders of Sinsoma GmbH. a for
413 profit company dedicated to the analysis of environmental DNA.
414

References

Agersnap, S., Larsen, W. B., Knudsen, S. W., Strand, D., Thomsen, P. F., Hesselsøe, M., … Møller, P. R. (2017). Monitoring of noble, signal and narrow-clawed crayfish using environmental DNA from freshwater samples. *PLOS ONE*, *12*(6), e0179261. doi:10.1371/journal.pone.0179261

Barnes, M. A., & Turner, C. R. (2016). The ecology of environmental DNA and implications for conservation genetics. *Conservation Genetics*, *17*(1), 1–17. doi:10.1007/s10592-015-0775-4

Biggs, J., Ewald, N., Valentini, A., Gaboriaud, C., Dejean, T., Griffiths, R. A., … Dunn, F. (2015). Using eDNA to develop a national citizen science-based monitoring programme for the great crested newt (*Triturus cristatus*). *Biological Conservation*, *183*, 19–28. doi:10.1016/j.biocon.2014.11.029

Blackman, R., Mächler, E., Altermatt, F., Arnold, A., Beja, P., Boets, P., … Deiner, K. (2019). Advancing the use of molecular methods for routine freshwater macroinvertebrate biomonitoring – the need for calibration experiments. *Metabarcoding and Metagenomics*, *3*, e34735. doi:10.3897/mbmg.3.34735

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., & Coissac, E. (2016). obitools: A unix-inspired software package for DNA metabarcoding. *Molecular Ecology Resources*, *16*(1), 176–182. doi:10.1111/1755-0998.12428

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*(4), 611–622. doi:10.1373/clinchem.2008.112797

Buxton, A. S., Groombridge, J., & Griffiths, R. (2018a). Comparison of Two Citizen Scientist Methods for Collecting Pond Water Samples for Environmental DNA Studies. *Citizen Science: Theory and Practice*, *3*(2). doi:10.5334/cstp.151

441 Buxton, A. S., Groombridge, J. J., & Griffiths, R. A. (2017). Is the detection of aquatic
442 environmental DNA influenced by substrate type? *PLOS ONE*, *12*(8), e0183371.
443 doi:10.1371/journal.pone.0183371
444 Buxton, A. S., Groombridge, J. J., & Griffiths, R. A. (2018b). Seasonal variation in environmental
445 DNA detection in sediment and water samples. *PLOS ONE*, *13*(1).
446 doi:10.1371/journal.pone.0191737
447 Buxton, A. S., Groombridge, J. J., Zakaria, N. B., & Griffiths, R. A. (2017, April 10). Seasonal
448 variation in environmental DNA in relation to population size and environmental factors.
449 *Scientific Reports*. Nature Publishing Group. doi:10.1038/srep46294
450 Cannon, M. V., Hester, J., Shalkhauser, A., Chan, E. R., Logue, K., Small, S. T., & Serre, D.
451 (2016). In silico assessment of primers for eDNA studies using PrimerTree and application
452 to characterize the biodiversity surrounding the Cuyahoga River. *Scientific Reports*, *6*.
453 doi:10.1038/srep22908
454 De'Ath, G., & Fabricius, K. E. (2000). Classification and regression trees: A powerful yet simple
455 technique for ecological data analysis. *Ecology*, *81*(11), 3178–3192. doi:10.1890/0012-
456 9658(2000)081[3178:CARTAP]2.0.CO;2
457 de Souza, L. S., Godwin, J. C., Renshaw, M. A., & Larson, E. (2016). Environmental DNA
458 (eDNA) detection probability is influenced by seasonal activity of organisms. *PLOS ONE*,
459 *11*(10), e0165273. doi:10.1371/journal.pone.0165273
460 Deiner, K., & Altermatt, F. (2014). Transport distance of invertebrate environmental DNA in a
461 natural river. *PLOS ONE*, *9*(2). doi:10.1371/journal.pone.0088786
462 Deiner, K., Bik, H. M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., …
463 Bernatchez, L. (2017). Environmental DNA metabarcoding: Transforming how we survey
464 animal and plant communities. *Molecular Ecology*, *26*(21), 5872–5895.
465 doi:10.1111/mec.14350
466 Doi, H., Katano, I., Sakata, Y., Souma, R., Kosuge, T., Nagano, M., … Tojo, K. (2017).

Detection of an endangered aquatic heteropteran using environmental DNA in a wetland ecosystem. *Royal Society Open Science*, *4*(7). doi:10.1098/rsos.170568

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*(9), 5601–5608. doi: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*(3), 1–11. doi:10.1371/journal.pone.0122763

Dorazio, R. M., & Erickson, R. A. (2018). ednaoccupancy□: An r package for multiscale occupancy modelling of environmental DNA data. *Molecular Ecology Resources*, *18*(2), 368–380. doi:10.1111/1755-0998.12735

Dunn, N., Priestley, V., Herraiz, A., Arnold, R., & Savolainen, V. (2017). Behavior and season affect crayfish detection and density inference using environmental DNA. *Ecology and Evolution*, *7*(19), 7777–7785. doi:10.1002/ece3.3316

Elbrecht, V., & Leese, F. (2017). PrimerMiner: an R package for development and in silico validation of DNA metabarcoding primers. *Methods in Ecology and Evolution*, *8*(5), 622–626. doi:10.1111/2041-210X.12687

Furlan, E. M., & Gleeson, D. (2017). Improving reliability in environmental DNA detection surveys through enhanced quality control. *Marine and Freshwater Research*, *68*(2), 388. doi:10.1071/MF15349

Goldberg, C. S., Strickler, K. M., & Fremier, A. K. (2018). Degradation and dispersion limit environmental DNA detection of rare amphibians in wetlands: Increasing efficacy of sampling designs. *Science of the Total Environment*, *633*, 695–703. doi:10.1016/j.scitotenv.2018.02.295

Goldberg, C. S., Turner, C. R., Deiner, K., Klymus, K. E., Thomsen, P. F., Murphy, M. A., …

493 Taberlet, P. (2016). Critical considerations for the application of environmental DNA
494 methods to detect aquatic species. *Methods in Ecology and Evolution*, *7*(11), 1299–1307.
495 doi:10.1111/2041-210X.12595
496 Halling, K. C., Schrijver, I., & Persons, D. L. (2012). Test verification and validation for molecular
497 diagnostic assays. *Archives of Pathology and Laboratory Medicine*, *136*(1), 11–13.
498 doi:10.5858/arpa.2011-0212-ED
499 Harper, K. J., Anucha, N. P., Turnbull, J. F., Bean, C. W., & Leaver, M. J. (2018). Searching for
500 a signal: Environmental DNA (eDNA) for the detection of invasive signal crayfish,
501 *Pacifastacus leniusculus* (Dana, 1852). *Management of Biological Invasions*, *9*(2), 137–
502 148. doi:10.3391/mbi.2018.9.2.07
503 Harper, L. R., Buxton, A. S., Rees, H. C., Bruce, K., Brys, R., Halfmaerten, D., … Hänfling, B.
504 (2019). Prospects and challenges of environmental DNA (eDNA) monitoring in freshwater
505 ponds. *Hydrobiologia*, *826*(1), 25–41. doi:10.1007/s10750-018-3750-5
506 Harper, L. R., Griffiths, N. P., Lawson Handley, L., Sayer, C. D., Read, D. S., Harper, K. J., …
507 Hänfling, B. (2019). Development and application of environmental DNA surveillance for
508 the threatened crucian carp (*Carassius carassius*). *Freshwater Biology*, *64*(1), 93–107.
509 doi:10.1111/fwb.13197
510 Hunter, M. E., Dorazio, R. M., Butterfield, J. S. S., Meigs-Friend, G., Nico, L. G., & Ferrante, J.
511 A. (2017). Detection limits of quantitative and digital PCR assays and their influence in
512 presence–absence surveys of environmental DNA. *Molecular Ecology Resources*, *17*(2),
513 221–229. doi:10.1111/1755-0998.12619
514 Hunter, M. E., Oyler-McCance, S. J., Dorazio, R. M., Fike, J. A., Smith, B. J., Hunter, C. T., …
515 Hart, K. M. (2015). Environmental DNA (eDNA) sampling improves occurrence and
516 detection estimates of invasive Burmese pythons. *PLOS ONE*, *10*(4), e0121655.
517 doi:10.1371/journal.pone.0121655
518 Jerde, C. L. (2019). Can we manage fisheries with the inherent uncertainty from eDNA? *Journal*

519 *of Fish Biology*, jfb.14218. doi:10.1111/jfb.14218

520 Klymus, K. E., Merkes, C. M., Allison, M. J., Goldberg, C. S., Helbing, C. C., Hunter, M. E., …
521 Richter, C. A. (2019). Reporting the limits of detection and quantification for environmental
522 DNA assays. *Environmental DNA*, edn3.29. doi:10.1002/edn3.29

523 Lacoursière-Roussel, A., & Deiner, K. (2019). Environmental DNA is not the tool by itself.
524 *Journal of Fish Biology*, jfb.14177. doi:10.1111/jfb.14177

525 Larson, E. R., Renshaw, M. A., Gantz, C. A., Umek, J., Chandra, S., Lodge, D. M., & Egan, S.
526 P. (2017). Environmental DNA (eDNA) detects the invasive crayfishes *Orconectes rusticus*
527 and *Pacifastacus leniusculus* in large lakes of North America. *Hydrobiologia*, *800*(1), 173–
528 185. doi:10.1007/s10750-017-3210-7

529 Lawson Handley, L., Read, D. S., Winfield, I. J., Kimbell, H., Johnson, H., Li, J., … Hänfling, B.
530 (2019). Temporal and spatial variation in distribution of fish environmental DNA in
531 England's largest lake. *Environmental DNA*, *1*(1), 26–39. doi:10.1002/edn3.5

532 Leese, F., Altermatt, F., Bouchez, A., Ekrem, T., Hering, D., Meissner, K., … Zimmermann, J.
533 (2016). DNAqua-Net: Developing new genetic tools for bioassessment and monitoring of
534 aquatic ecosystems in Europe. *Research Ideas and Outcomes*, *2*, e11321.
535 doi:10.3897/rio.2.e11321

536 MacDonald, A. J., & Sarre, S. D. (2017). A framework for developing and validating taxon-
537 specific primers for specimen identification from environmental DNA. *Molecular Ecology*
538 *Resources*, *17*(4), 708–720. doi:10.1111/1755-0998.12618

539 MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, A., & Langtimm, C. A.
540 (2002). Estimating site occupancy rates when detection probabilities are less than one.
541 *Ecology*, *83*(8), 2248–2255.

542 Mauvisseau, Q., Coignet, A., Delaunay, C., Pinet, F., Bouchon, D., & Souty-Grosset, C. (2018).
543 Environmental DNA as an efficient tool for detecting invasive crayfishes in freshwater
544 ponds. *Hydrobiologia*, *805*(1), 163–175. doi:10.1007/s10750-017-3288-y

Meusnier, I., Singer, G. A. C., Landry, J. F., Hickey, D. A., Hebert, P. D. N., & Hajibabaei, M. (2008). A universal DNA mini-barcode for biodiversity analysis. *BMC Genomics*, *9*. doi:10.1186/1471-2164-9-214

Natural England. (2015). Great crested newts: protection and licences. Retrieved 7 April 2020, from https://www.gov.uk/guidance/great-crested-newts-protection-surveys-and-licences

Nicholson, A., McIsaac, D., MacDonald, C., Gec, P., Mason, B. E., Rein, W., … Hanner, R. H. (2020). An analysis of metadata reporting in freshwater environmental DNA research calls for the development of best practice guidelines. *Environmental DNA*. doi:10.1002/edn3.81

R Core Team. (2019). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Rees, H. C., Bishop, K., Middleditch, D. J., Patmore, J. R. M., Maddison, B. C., & Gough, K. C. (2014). The application of eDNA for monitoring of the Great Crested Newt in the UK. *Ecology and Evolution*, *4*(21), 4023–4032. doi:10.1002/ece3.1272

Rees, H. C., Maddison, B. C., Middleditch, D. J., Patmore, J. R. M., & Gough, K. C. (2014). The detection of aquatic animal species using environmental DNA - a review of eDNA as a survey tool in ecology. *Journal of Applied Ecology*, *51*(5), 1450–1459. doi:10.1111/1365-2664.12306

Robinson, C. V., Uren Webster, T. M., Cable, J., James, J., & Consuegra, S. (2018). Simultaneous detection of invasive signal crayfish, endangered white-clawed crayfish and the crayfish plague pathogen using environmental DNA. *Biological Conservation*, *222*, 241–252. doi:10.1016/j.biocon.2018.04.009

Thalinger, B., Oehm, J., Mayr, H., Obwexer, A., Zeisler, C., & Traugott, M. (2016). Molecular prey identification in Central European piscivores. *Molecular Ecology Resources*, *16*(1), 123–137. doi:10.1111/1755-0998.12436

Thalinger, B., Wolf, E., Traugott, M., & Wanzenböck, J. (2019). Monitoring spawning migrations of potamodromous fish species via eDNA. *Scientific Reports*, *9*(1). doi:10.1038/s41598-

571 019-51398-0

572 Thomsen, P. F., Kielgast, J., Iversen, L. L., Wiuf, C., Rasmussen, M., Gilbert, M. T. P., …
573 Willerslev, E. (2012). Monitoring endangered freshwater biodiversity using environmental
574 DNA. *Molecular Ecology*, *21*(11), 2565–2573. doi:10.1111/j.1365-294X.2011.05418.x

575 Wilcox, T. M., Carim, K. J., McKelvey, K. S., Young, M. K., & Schwartz, M. K. (2015). The dual
576 challenges of generality and specificity when developing environmental DNA markers for
577 species and subspecies of *Oncorhynchus*. *PLOS ONE*, *10*(11), e0142008.
578 doi:10.1371/journal.pone.0142008

579 Wilcox, T. M., Young, M. K., McKelvey, K. S., Isaak, D. J., Horan, D. L., & Schwartz, M. K.
580 (2018). Fine-scale environmental DNA sampling reveals climate-mediated interactions
581 between native and invasive trout species. *Ecosphere*, *9*(11), e02500.
582 doi:10.1002/ecs2.2500

583 Williams, M. A., O'Grady, J., Ball, B., Carlsson, J., de Eyto, E., McGinnity, P., … Parle-
584 McDermott, A. (2019). The application of CRISPR-Cas for single species identification from
585 environmental DNA. *Molecular Ecology Resources*, *19*(5), 1106–1114. doi:10.1111/1755-
586 0998.13045

587 Williams, M. R., Stedtfeld, R. D., Engle, C., Salach, P., Fakher, U., Stedtfeld, T., … Hashsham,
588 S. A. (2017). Isothermal amplification of environmental DNA (eDNA) for direct field-based
589 monitoring and laboratory confirmation of *Dreissena* sp. *PLOS ONE*, *12*(10).
590 doi:10.1371/journal.pone.0186462

591 World Health Organization. (2011). *Laboratory quality standards and their implementation.* New
592 Delhi: WHO Regional Office for the Western Pacific.

593 Yamanaka, H., & Minamoto, T. (2016). The use of environmental DNA of fishes as an efficient
594 method of determining habitat connectivity. *Ecological Indicators*, *62*, 147–153.
595 doi:10.1016/j.ecolind.2015.11.022

596 Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). Primer-

BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, *13*, 134. doi:10.1186/1471-2105-13-134

Tables and Figures

Table 1: A guideline for troubleshooting at different stages of the validation process.

Table 2: Detection methods and their trade-offs used for targeted eDNA assays. Abbreviations are as follows: polymerase chain reaction (PCR), capillary electrophoresis (CE), quantitative (q), intercalating dye (ID), loop-mediated isothermal amplification (LAMP).

Table 3: The thematic variable blocks of the 5-level validation scale and their respective minimum criteria.

Figure 1: An overview of the 5-level validation scale. For each of the levels (incomplete to operational), the main accomplishments in the validation process and an appropriate interpretation of results are provided.

Figure 2: The main outcomes of the meta-analysis based on 546 assays from 327 publications are presented in panels A to E. Assay classification is based on the minimum criteria presented in Figure 1. Level 0 codes for assays that did not reach Level 1 on the validation scale. The colour coding is consistent for all panels: Level 0 (grey), Level 1 (dark purple), Level 2 (blue), Level 3 (turquoise), Level 4 (green), and Level 5 (yellow). Panel A shows the distribution of assays across levels of the validation scale. Panel B displays the percentage of assays (*N* = 546) rated Level 0 to 4 that have been published each year since 2003. Panel C summarises variable reporting per assay level. Panel D shows the percentage of assays reporting a specific variable (colour-coded according to level). Panel E shows the minimum criteria necessary to reach each level of the validation scale, and the percentage of Level 0 to 4 assays that did not report these. All variable abbreviations are listed in SI1.

626

627 **Figure 3:** Classification tree analysis identifying the criteria distinguishing assays at different
628 levels of the validation scale. The conditions along the branches show the criteria on which the
629 dataset is split. Numbers in coloured leaves show the validation level of the assays in the
630 respective leaf. Numbers below the leaves represent the number of assays per validation level,
631 summarised inside an individual leaf. The displayed percentage is the proportion of assays
632 summarised per leaf.

633

634 **Box 1:** Examples of assays rated at Levels 0 to 4. The vertical tile plot shows which of the
635 minimum criteria the assay fulfils (yellow tiles), and the bar chart gives the scoring percentage
636 (i.e. the proportion of variables that were tested or reported) for each of the variable blocks.
637 Bars are coloured according to the score obtained from a block with dark purple coding for “no
638 validation” and yellow coding for “comprehensive validation”.

Table 1:

General issues	*In silico* validation	*In vitro* validation	*In situ* validation
Applying assay to a new geographic location	Test non-target sequences available on public/custom databases from this location	Test tissue from non-target species present at this location, including species present at original location if populations are genetically diverse	
Sparse reference data	Where possible, Sanger sequence tissue DNA from target and non-target species to generate reference sequences	Extensively test assay on tissue available for target and non-target species, i.e. test extracts from multiple individuals	Extensively test assay at sites where target species co-occurs with non-target species lacking reference sequences, and at sites where only non-target species lacking reference sequences occur

In vitro validation issues	Troubleshooting guidelines	*In situ* validation issues	Troubleshooting guidelines
Poor amplification efficiency	Optimise reaction volume	**Expected species presence/absence not confirmed**	Survey sites with conventional tools and eDNA metabarcoding if possible
	Optimise cycle number		
	Optimise primer-probe concentration	**False positives**	Discard samples corresponding to contaminated controls; consider lack of specificity; Sanger sequence results
	Optimise annealing temperature using gradient PCR		
	Optimise technical replication		Resample with more stringent decontamination measures in place
	Check pipetting accuracy		
Poor specificity	Increase annealing temperature	**False negatives**	Check for inhibition and treat samples if inhibited; consider timing and spacing of sampling
	Add hydrolysis probe		
	Perform melting curve analysis		Increase technical replication
	Consider redesign		Resample and increase biological replication and volume of water collected
Low sensitivity (see also Klymus et al., 2019)	Ensure enough replication of fresh standards used when establishing the Limit of Detection	**Poor quality Sanger sequencing**	Purify amplicons prior to sequencing
	Use TE buffer and tRNA to make standard dilutions, not molecular grade water		Concentrate amplicons prior to sequencing
	Use low retention tubes and pipette tips when preparing standards to prevent adsorption to plastic	**Inhibition (determined by failed/skewed amplification of IPC or non-target assay)**	Use a DNA extraction kit that includes an inhibitor removal step (e.g. mu-DNA, Qiagen QIAamp DNA Stool Mini Kit, Qiagen DNeasy PowerWater Kit, Qiagen DNeasy PowerSoil Kit)
	If possible, switch to PCR platform with greater sensitivity		Apply an inhibitor removal kit (e.g. Zymo) to samples
	Consider redesign		Use PCR reagents designed to handle inhibition (e.g. TaqMan Environmental Master Mix, Bovine Serum Albumin)

Table 2:

Trade-offs	Endpoint PCR-gel	Endpoint PCR-CE	qPCR-ID	qPCR-probe	dPCR- ID	dPCR-probe	LAMP
Quantification	no	yes	yes	yes	yes	yes	yes, with real-time monitoring
LOD (sensitivity)	medium - high DNA concentration (depending on which intercalating dye)	low - medium DNA concentration (with a well designed assay)	low DNA quantities	low DNA quantities	absolute DNA quantities	absolute DNA quantities	medium DNA concentration
Specificity	medium	medium	medium	high	medium	high	high
Cost reagents	low	medium	high	high	high	very high	medium
Cost equipment	low	medium	medium	medium	high	high	low - medium (depending on LAMP cycler)
Multiplex possible	yes, if different sizes are used	yes, if different sizes or dyes are used	difficult	yes, if different dyes are used	difficult	yes, if different dyes are used	yes
Time (PCR to data)	slow	slow	medium	medium	slow	slow	fast (15-30 min)
Assay type transferable	yes	yes	yes	yes	yes to qPCR	yes to qPCR	not easily
Effort to design assay	low - medium	low - medium	high	high	high	high	high
Specialised software	advisable	advisable	advisable	required	advisable	required	required
Sequencing confirmation	yes	yes	yes	yes	limited	limited	limited

Table 3:

Validation level	Variable blocks	Minimum criteria
Level 1	*in silico* analysis	target species
	target tissue testing	target tissue
	target tissue PCR	primer (and probe) sequence
Level 2	comprehensive reporting of PCR conditions	DNA extract volume in PCR
	in vitro testing on closely related non-target species	any in vitro non-target testing
Level 3	extraction method performed on eDNA samples	method of extraction
	concentration of eDNA from environmental sample	filter type or precipitation chemicals
	detection obtained from environmental samples	detection from an environmental sample (artificial or natural habitat)
Level 4	Limit of Detection (LOD)	LOD determined
	extensive field testing of environmental samples	multiple locations or multiple samples
	in vitro testing on co-occurring non-target species	any advanced *in vitro* testing
Level 5	comprehensive specificity testing	non-co-occurring/closely related species checked from *in silico*
	detection probability estimation from statistical modelling	any effort made towards detection probability estimation
	understanding ecological and physical factors influencing eDNA in the environment	any factor influencing eDNA in the environment tested

Figure 1

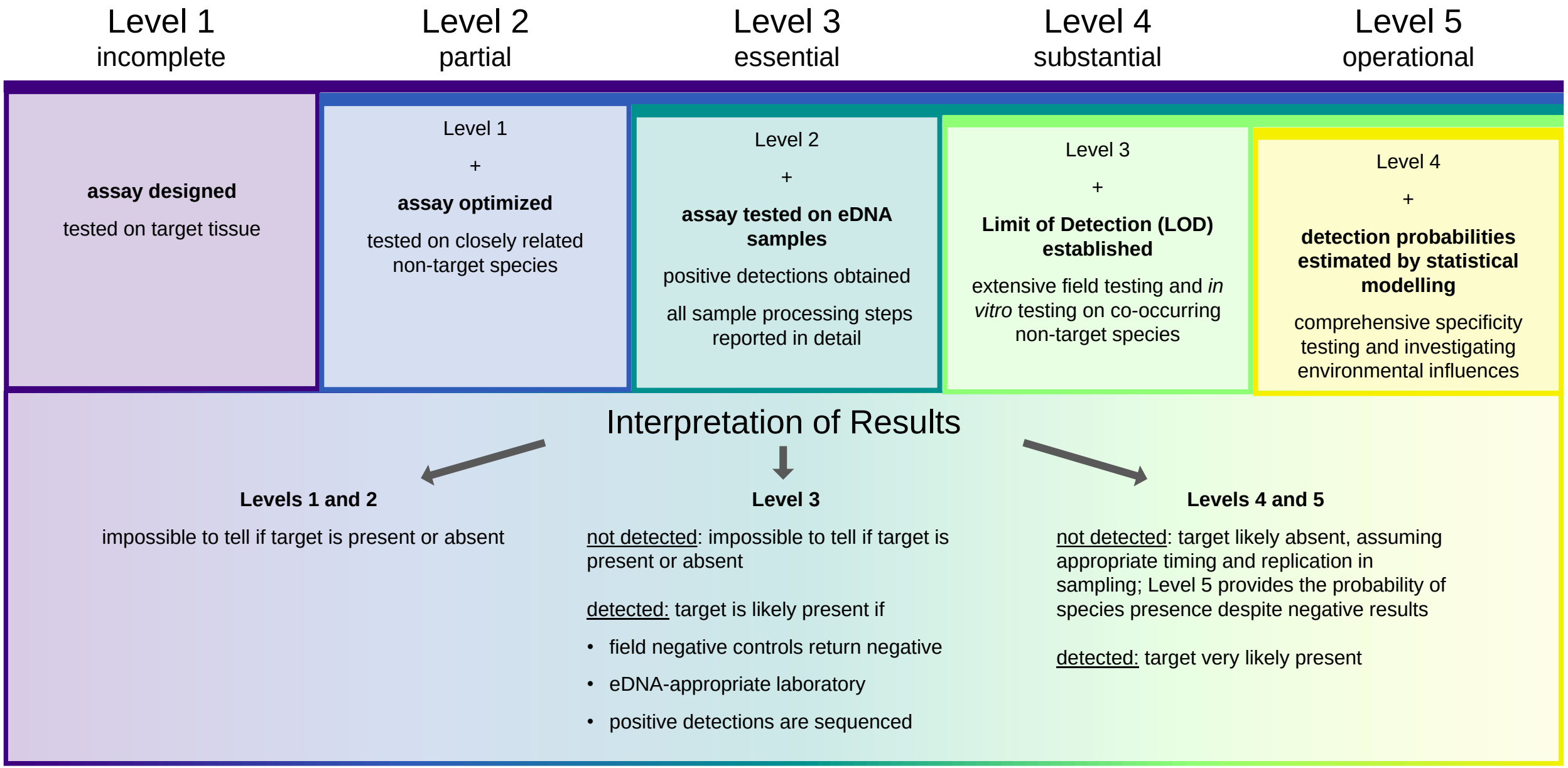

Level 1
incomplete
Level 2
partial
Level 3
essential
Level 4
substantial
Level 5
operational
assay designed
tested on target tissue
Level 1
+
assay optimized
tested on closely related non-target species
Level 2
+
assay tested on eDNA samples
positive detections obtained
all sample processing steps reported in detail
Level 3
+
Limit of Detection (LOD) established
extensive field testing and in vitro testing on co-occurring non-target species
Level 4
+
detection probabilities estimated by statistical modelling
comprehensive specificity testing and investigating environmental influences
Interpretation of Results
Levels 1 and 2
impossible to tell if target is present or absent
Level 3
not detected: impossible to tell if target is present or absent
detected: target is likely present if
• field negative controls return negative
• eDNA-appropriate laboratory
• positive detections are sequenced
Levels 4 and 5
not detected: target likely absent, assuming appropriate timing and replication in sampling; Level 5 provides the probability of species presence despite negative results
detected: target very likely present

Figure 2:

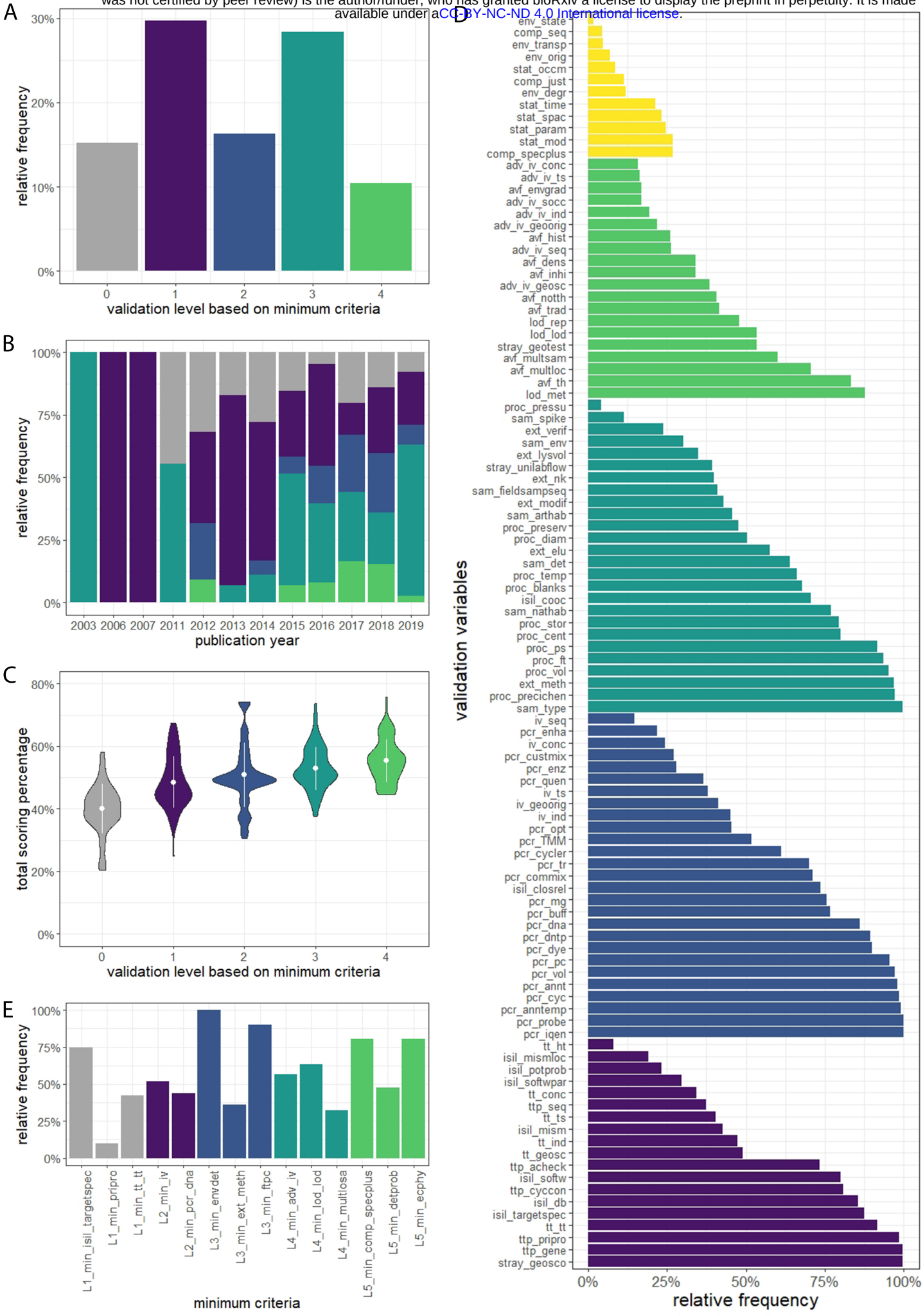
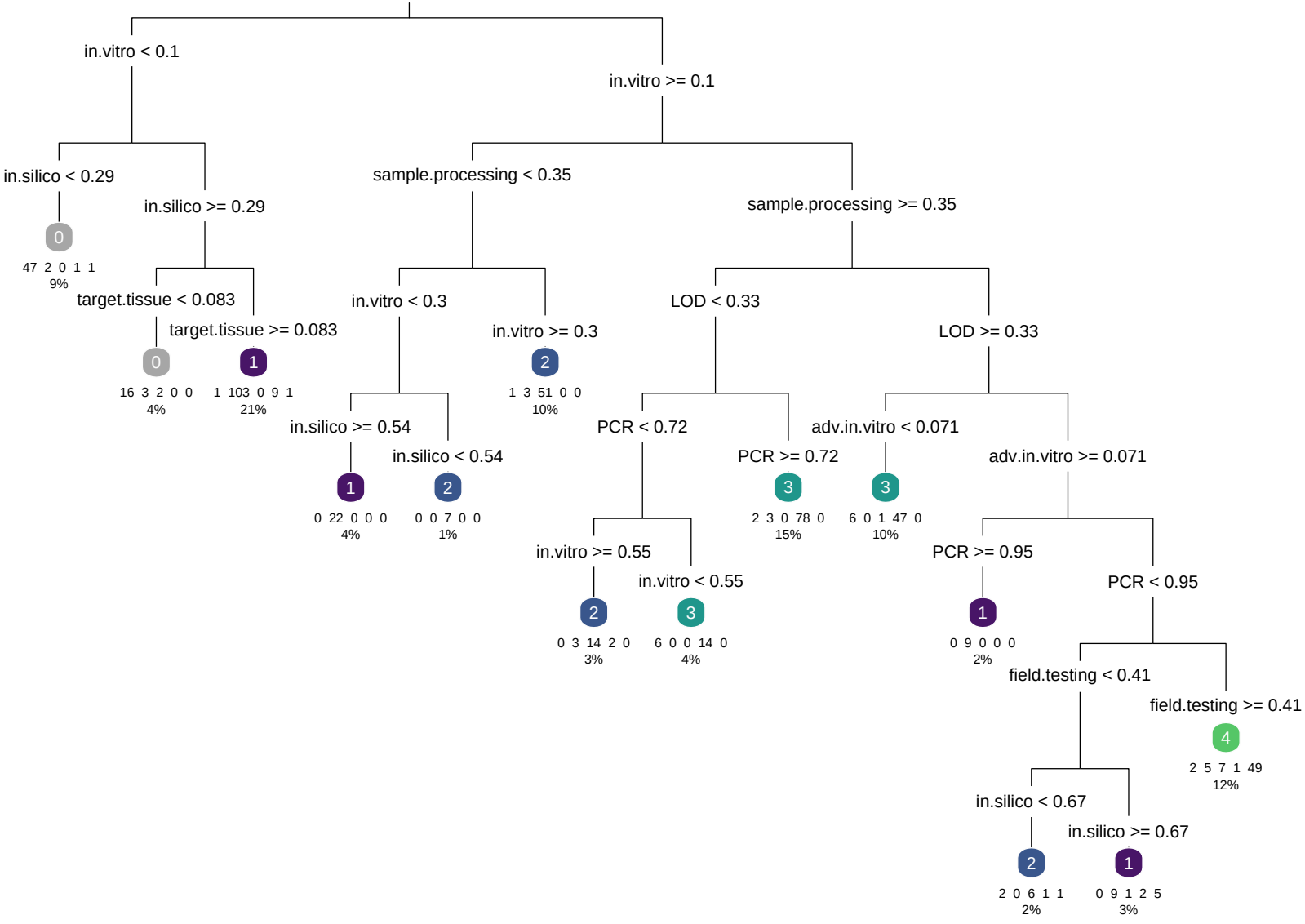
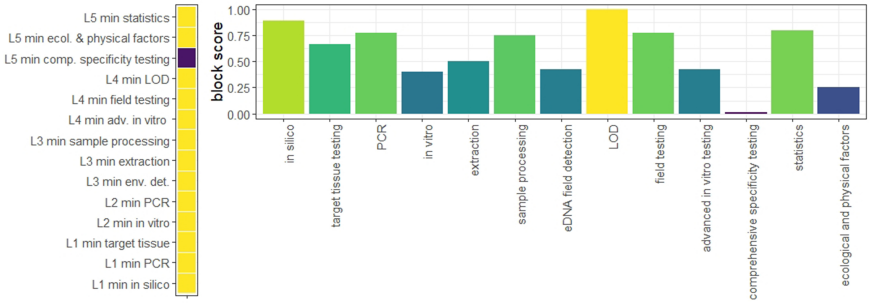


Figure 3:

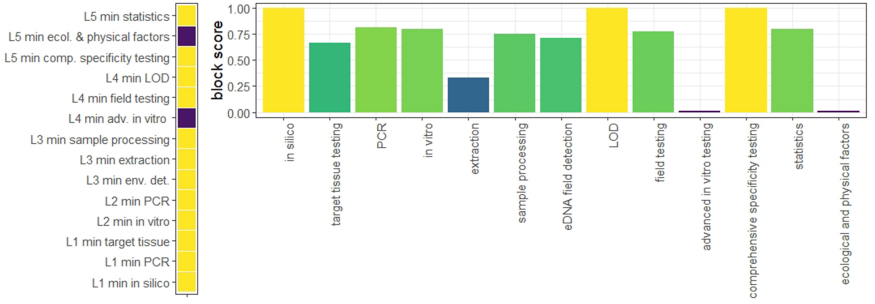

in.vitro < 0.1
in.vitro >= 0.1
in.silico < 0.29
in.silico >= 0.29
sample.processing < 0.35
sample.processing >= 0.35
0
47 2 0 1 1
9%
target.tissue < 0.083
target.tissue >= 0.083
in.vitro < 0.3
in.vitro >= 0.3
LOD < 0.33
LOD >= 0.33
0
16 3 2 0 0
4%
1
1 103 0 9 1
21%
2
1 3 51 0 0
10%
in.silico >= 0.54
in.silico < 0.54
PCR < 0.72
PCR >= 0.72
adv.in.vitro < 0.071
adv.in.vitro >= 0.071
1
0 22 0 0 0
4%
2
0 0 7 0 0
1%
3
2 3 0 78 0
15%
3
6 0 1 47 0
10%
in.vitro >= 0.55
in.vitro < 0.55
PCR >= 0.95
PCR < 0.95
2
0 3 14 2 0
3%
3
6 0 0 14 0
4%
1
0 9 0 0 0
2%
field.testing < 0.41
field.testing >= 0.41
4
2 5 7 1 49
12%
in.silico < 0.67
in.silico >= 0.67
2
2 0 6 1 1
2%
1
0 9 1 2 5
3%

Box 1:

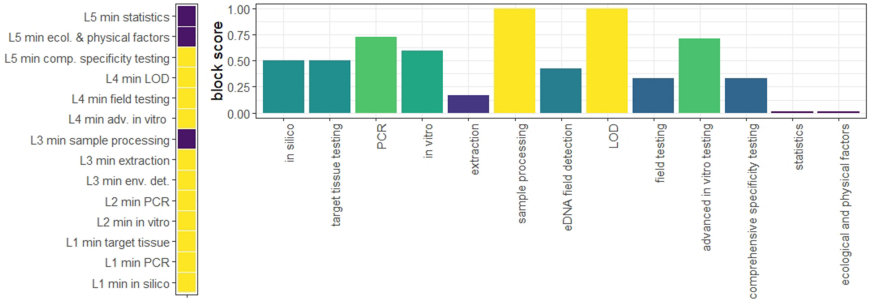
Level 4: Assay no. 165 detecting brook trout, *Salvelinus fontinalis*. Assay information was published in paper nos. 291, 12, 145, 295, 144, with latest publication in 2018. Despite low validation efforts for *in vitro* testing and field detection, the only minimum criterion that prevented this assay attaining Level 5 was “comprehensive specificity testing”.



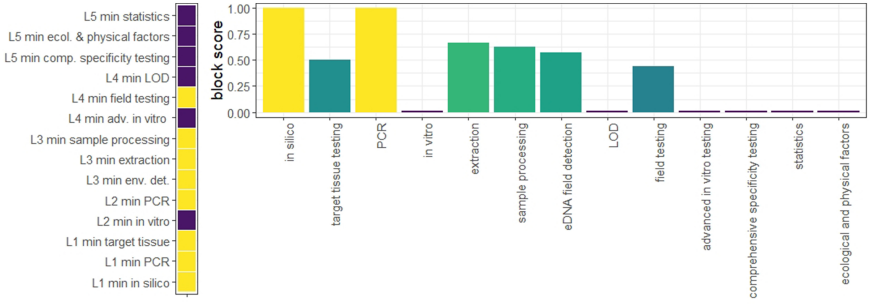
Level 3: Assay no. 473 detecting the crucian carp, *Carassius carassius*. Assay information was published in paper no. 119 in 2019. No advanced *in vitro* testing was conducted for this assay, but validation efforts were substantial for most thematic blocks, especially “*in silico* testing”, “LOD determination”, and “comprehensive specificity testing”.



Level 2: Assay no. 270 detecting common eelgrass, *Zostera marina*. Assay information was published in paper no. 117 in 2018. Despite low validation efforts in many thematic blocks, only the minimum criterion of “sample processing” prevented this assay from reaching Level 4 on the validation scale



Level 1: Assay no. 315 detecting the waterlouse, *Asellus aquaticus*. Assay information was published in paper no. 163 in 2015. With additional *in vitro* testing, this assay would have attained Level 3. A clear decrease in validation efforts is visible for variable blocks associated with higher levels of the validation scale.



Level 0: Assay no. 343 detecting *Daphnia magna*. Assay information was published in paper no. 243 in 2018. Despite some advanced validation, this assay does not fulfill the minimum criterion of “*in silico* testing” and “target tissue testing”. Therefore, it could not be classed as Level 1 on the validation scale.