Minimum Performance Parameters

First version. **Deliverable 6.1**

PROJECT NUMBER: FP7-KBBE-2013-7-613908-Decathlon

Scope of this guidance document

The present guidance document is intended to facilitate establishment of quality assuring, cross-sectorial, reasonable and sufficiently strict performance requirement for development, validation and application of molecular analytical methods. The present guidance document is the first version of what will become an amended and final version at the end of 2016. The present version is limited both in number of specific performance parameters and in its sectorial basis. The final version will include many more performance parameters and have a much broader sectorial basis. The present version will hopefully stimulate representatives of multiple sectors where molecular analytical methods are applicable, to contribute to the realization of a truly cross-sectorial guidance document adding valuable information to all relevant stakeholders and facilitating harmonization.

Authors

The document was prepared collectively by the Decathlon project partners, under the umbrella of work package 6 “Minimum performance parameters (MPP) and validation aspects”. This work package is coordinated by Arne Holst-Jensen from the Norwegian Veterinary Institute, who is also task leader for the subsection on next generation sequencing. He is assisted by task leaders for subsections relating to DNA extraction (Teresa Crespo and Ana Simplicio, Instituto de Biologia Experimental Tecnologica, Portugal), PCR amplification (Petra Richl and Mathias Welsche, Eurofins GeneScan, Germany), and isothermal amplification (David Dobnik and Tanja Dreo, National Institute of Biology, Slovenia). The many contributions from colleagues and project partners (none named, none forgotten) have been extremely useful in the preparation of this first version of the MPP guidance document.

The Decathlon project has been funded with support from the European Commission in the context of the Seventh Framework Programme (FP7). This publication and all its contents reflect the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Table of Contents

1. Minimum Performance Parameters (MPPs) for DNA extraction methods, PCR and isothermal amplification methods and next generation sequencing (NGS) methods 4

1.1. Minimum performance parameters and associated acceptance values 4

1.2. The role of Decathlon and external stakeholders 5

1.3. Evolution of analytical modules, methods and this guideline document 5

2. The application of MPPs and AAV to the whole method 8

3. Key MPPs and AAVs included in present document: 10

3.1. Applicability 10

3.2. Specificity 10

3.3. False positive rate (type I error rate) 12

3.4. False negative rate (type II error rate) 13

3.5. Limit of detection (LOD) 13

3.6. Lower limit of quantification (LOQ) 15

3.7. Dynamic range 15

3.8. Accuracy (trueness and precision) 16

3.9. Trueness 18

3.10. Precision 18

3.11. Probability of Detection (POD) 20

4. Glossary: 21

5. References 24

5.1. Papers cited in the report 24

5.2. Papers consulted but not cited specifically 25

1. Minimum Performance Parameters (MPPs) for DNA extraction methods, PCR and isothermal amplification methods and next generation sequencing (NGS) methods

1.1. Minimum performance parameters and associated acceptance values

A minimum performance parameter (MPP) is a checkpoint for an analytical module (see chapter 4 on glossary for explanation of distinction between method and module). An associated acceptance value (AAV) is the critical (threshold) value for a specific MPP that the performance of the module must comply to. By defining the MPPs and AAVs upfront to development of a module for a specific purpose, and prior to validation and application of the module, it is expected that the general quality of molecular analytical tools will increase and that underperforming modules will be less common and more rapidly replaced by better modules. A module can then be benchmarked against these MPPs and AAVs. Failure to meet the AAV for a specific MPP means that the module is underperforming on this particular quality criterion. It is possible to develop a module without consideration of MPP and AAV, and then assess its performance. This, however, can reduce the chance of developing a module that is fit for successive use. Such approaches are not recommendable and will not be further discussed in the present guidance document.

MPPs are quality criteria. The MPPs and their AAVs can e.g. serve to ensure that modules and methods are compliant with regulatory requirements. A good example is the guidance document published by the European Union's Reference Laboratory for Genetically Modified Food and Feed in collaboration with the European Network of GMO Laboratories (EURL-GMFF 2008). The Decathlon project is funded by the European Commission (EC). It is reasonable to assume that the EC had regulatory requirements in mind when the call for proposal was launched and the project funded. However, we believe an equally important duty is to develop a document with broad, preferably global, cross sectorial applicability. This will require balancing of pragmatism and tough but feasible MPPs/AAVs. The definitions of MPPs must therefore be applicable across sectors, while the AAVs in some cases must be sector and even application specific. Much of the work laid down in preparation of the present guidance document is review based. Regulations, norms, scientific experience and

recommendations from experts from relevant sectors are the main information sources explored. In cases where no external requirements exist the AAVs of the present guideline document are set by the authors. These values are believed to be sufficiently pragmatic to be broadly applicable and sufficiently tough to avoid compromising analytical quality.

1.2. The role of Decathlon and external stakeholders

The Decathlon project is particularly focused on detection and identification of genetically modified organisms (GMOs), food pathogens and species and products subject to customs and excise duty regulations (see Decathlon website for details; www.decathlon-project.eu/). Given the much broader sectorial perspective of the present guidance document and later amendments (see the scope above), the input from external experts and stakeholders is absolutely essential. We therefore strongly encourage external experts and stakeholders, in particular from sectors other than those directly involved in Decathlon, to forward their views, suggestions, recommendations and experiences to http://www.decathlon-project.eu/content/contacts.

1.3. Evolution of analytical modules, methods and this guideline document

The life of analytical modules and methods can go through four phases: *i)* development; *ii)* validation; *iii)* routine application; and *iv)* substitution by alternative modules/methods (comparison, benchmarking and phasing out). Each MPP and AAV is not equally relevant for all phases (see Tables 1 and 2). In the following chapters we will highlight these differences.

In the first (developmental) phase a light set of MPPs should be sufficient to get a clear impression of the reliability of a module. Further development and validation should take place only if the module appears to be reliable, to minimize risk of wasting resources. The present document offers guidance to the assessment for each MPP. MPPs for phase *ii* are the most comprehensive and their role is to provide extensive evidence of the reliability of the module. Both the number of MPPs to be assessed and the number of samples, repetitions and participating laboratories usually exceed those of phase *i*. MPPs for phase *iii* are the controls intended to demonstrate that the execution of the module in combination with the materials subjected to analysis will provide reliable results.

Table 1. Minimum performance parameters (MPPs) included and when they are applied in the life history of an analytical module

MPP	Phase I (development)	Phase II (validation)	Phase III (routine application)
Applicability	X	X	X
Specificity	X	X	-
False positive rate	X	X	-
False negative rate	X	X	-
Limit of detection (LOD)	X	X	(X)[a]
Lower limit of quantification (LOQ)	X	X	(X)[a]
Dynamic range	X	X	-
Accuracy	X	X	X
Trueness	X	X	X
Precision	X	X	X
Intra laboratory precision (RSD_r)	X	X	X
Inter laboratory precision (RSD_R)	-	X	-
Probability of detection (POD)	-	X	-

[a] Practical limit of detection (pLOD) and practical limit of quantification (pLOQ) are very useful. These can often be determined and will then aid at defining the reliability of an analytical result and determine whether the analytical result meets the requirements of the end-user (e.g. if pLOD and pLOQ is inferior to a legal or contractual threshold or not).

Table 2. Minimum performance parameters (MPPs) included and where they are applied in relation to the whole method (see also Fig. 1)

MPP	Whole method	Sample prep., DNA extraction and purification	Nucleic acid amplification	Sequencing	Bioinformatics analysis
Applicability	X	X	X	X	X
Specificity	X	(X)[a]	X	(X)[d]	X
False positive rate	X	-	X	(X)[d]	X
False negative rate	X	-	X	(X)[d]	X
Limit of detection (LOD)	X	-	X	X	X
Lower limit of quantification (LOQ)	X	-	(X)[c]	(X)[e]	(X)[e]
Dynamic range	X	-	(X)[c]	(X)[e]	(X)[e]
Accuracy	X	-	(X)[c]	-	-
Trueness	X	(X)[b]	(X)[c]	-	-
Precision	X	-	(X)[c]	-	-
Probability of detection (POD)	X	-	X	X	X

[a] Specificity is important if selective enrichment is performed, e.g. using culturing, filtering or affinity based approaches.

[b] Bias can be introduced at this step, e.g. if the module introduces unequal probability of recovery of small vs. large DNA fragments or of AT rich vs. GC rich sequence motifs.

[c] Only applicable to quantitative modules.

[d] Sequencing is prone to errors from presence of non-target nucleic acid in the sample, and from mis-incorporation of label within or between samples.

[e] Sequencing is usually qualitative by nature, but in some cases the resulting data are applied and interpreted as semi-quantitative.

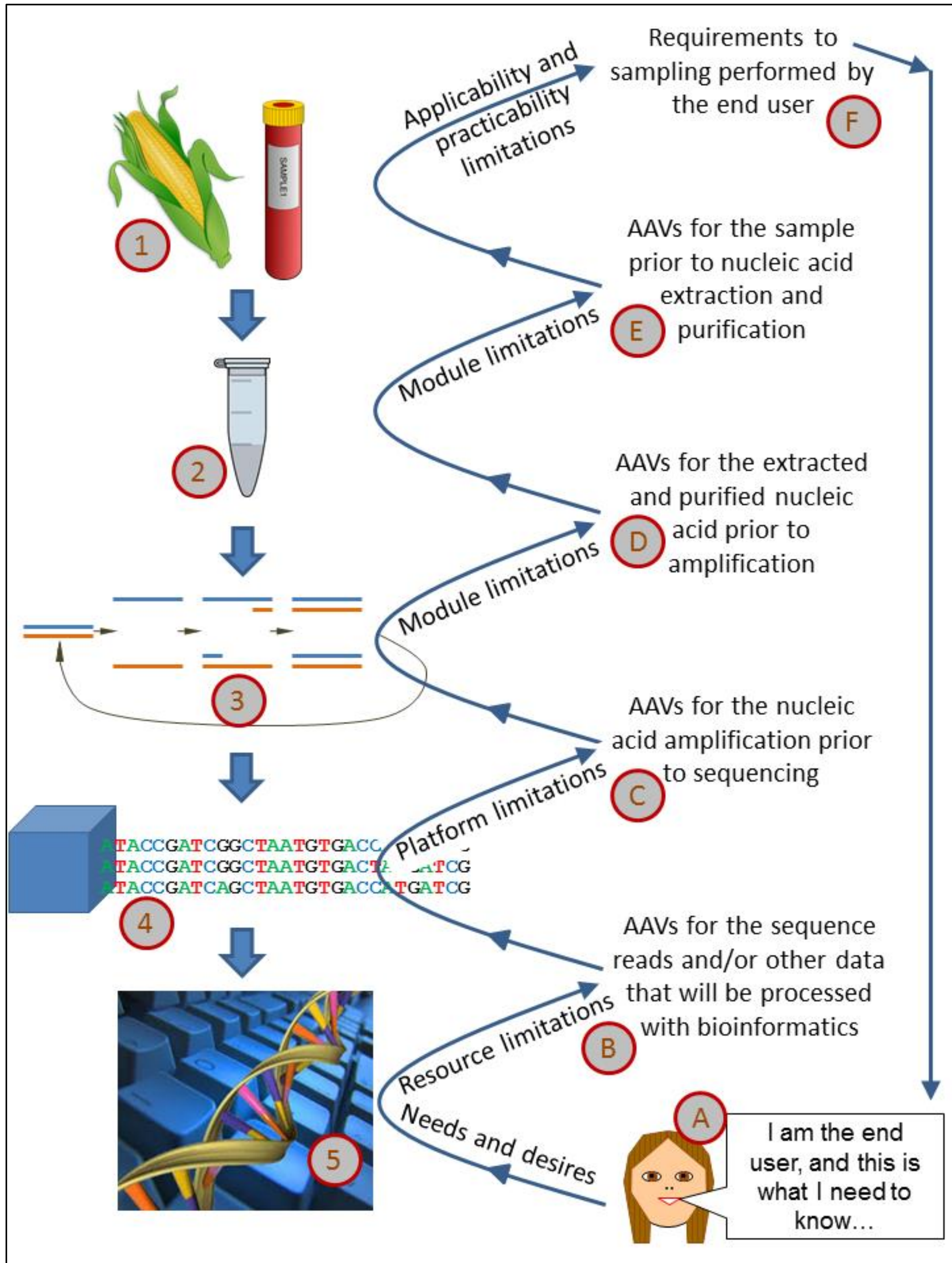
In phase *iii*, it can be necessary to consider the practical (sample specific) limit of detection (quantification) pLOD (pLOQ). For example in the case of GMO testing, it is not regulatory compliant in the EU to have a pLOD or pLOQ of 1.5% given that in the EU the labeling threshold is 0.9% (European Commission, 2003). Obviously the pLOD and pLOQ in this case must be < 0.9%. Thus, pLOD (pLOQ) is a relevant MPP, and it is necessary to make sure that the quantity and quality of the extracted DNA is sufficient to obtain a sufficiently good (low) pLOD (and/or pLOQ). The definition of "sufficiently good" is exactly the AAV in this case. The pLOD/pLOQ can also be an issue for other sectors, and potentially but not necessarily find its solution in quantifying the species DNA.

The present guidance document is the first version of an MPP guidance document for which a final amended version providing detailed guidance relating to DNA extraction, DNA amplification and next generation sequencing technologies will be made available in parallel to conclusion of the Decathlon project (end of 2016). The annex to the present guidance document provides an overview of MPPs and AAVs currently under consideration for the amendment. We warmly welcome external views and suggestions to this annex as it is our clear intention to have an open and inclusive dialogue with experts from a broad range of sectors. It has been and will continue to be a clear objective for the MPP guidance document to avoid sector specific bias of the MPPs and AAVs.

The list of MPPs includes some that will apply only to one phase or to one step of the whole analytical method, while others will apply to several phases or steps (see Tables 1 and 2). The project partners have agreed that for the MPPs for DNA extraction modules it is necessary to divide the MPPs in two groups to discriminate between the sectors and cases where pLOD (pLOQ) is relevant (quantitative PCR or equivalent tools must be used) and cases where pLOD (pLOQ) is irrelevant (more conventional approaches to assess quantity and quality can be applied). There is an ongoing discussion on how to best structure the MPPs relating to isothermal and PCR (including digital PCR) amplification modules. For Next Generation Sequencing most of the literature has a rather narrow set of quality control parameters, and we foresee the establishment of many more. An extensive list of candidates is presented in the annex which is available from http://www.decathlon-project.eu/article/download.

The context in which the whole method is applied has a strong influence on the weighting of MPPs. This is probably not sufficiently reflected in the present guidance document.

2. The application of MPPs and AAV to the whole method



← (previous page) **Figure 1. Top-down and bottom-up relationships between the whole method and the minimum performance parameters (MPPs) and associated acceptance values (AAVs).** The whole method is composed of a series of steps where the output of one step is the input to the next step (left). **1. → 2.** Sample preparation, nucleic acid extraction and purification. **2. → 3.** Nucleic acid processing, amplification and labeling. **3. → 4.** Sequencing of the nucleic acid. **4. → 5.** Bioinformatics, data processing and interpretation. The method shown here is a simplified illustration which includes amplification, sequencing and bioinformatics. Notably, one or more of these steps are optional, depending on the scope of the specific whole method. Furthermore, sequencing will typically involve library preparation which by itself may add steps to the diagram (not shown here). The quality and reliability of the final test result is dependent on acceptable performance at every step in this top-down chain. A bottom-up approach is taken when the modules forming a whole method are selected (right). **A. → B.** The end-user's needs should be taken as the starting point when the available bioinformatics resources and platforms are evaluated. **B. → C.** MPPs and AAVs influence on the applicability of available sequencing platforms. **C. → D.** MPPs and AAVs influence on the applicability of available optional sequencing library preparation approaches. **D. → E.** MPPs and AAVs influence on the applicability of sample preparation and nucleic acid extraction/purification modules. **E. → F.** MPPs and AAVs influence on the sampling, and these criteria must be communicated back to the end-user **(A)**. At each level in this chain the relevant limitations must be taken into consideration. Then the MPPs and AAVs are established for the preceding step (module level) of the whole method. This will establish a robust whole method where compliance to MPPs and AAVs at each step of the whole method will guarantee acceptable quality and reliability of the final test result.

3. Key MPPs and AAVs included in present document:

It is inevitable that some MPPs are linked with other MPPs. References to other MPPs are highlighted in ***bold italics*** in the following chapter. The following list of MPPs is organized in what we hope will be perceived as a logical structure. In this way we hope to facilitate reader's perception of contents compared to organizing the list alphabetically. However, as the list will grow considerably longer in the amendment, an alphabetical organization is foreseen in the next version of the MPP guidance document. An annex to the present document lists candidate MPPs pending inclusion (after substantial discussion) in the future amendment of the present document.

3.1. Applicability

Definition: description of the analytes, species, matrices and target concentrations to which the module can be applied, and the type of study/monitoring effort for which the module, judged from its performance, is suited.

Acceptance criterion: an applicability statement shall provide information on the scope of the module and include reference to data documenting the performance of the module. The description should also include warnings to known interferences by other analytes, or inapplicability to certain matrices and situations.

Note: The definition is modified from EURL-GMFF (2008) and ISO (2006).

Note: Applicability is sometimes referred to as fitness for purpose.

3.2. Specificity

Definition: Property of a method to respond exclusively to the characteristic or analyte of interest.

Acceptance criterion: the module should only produce a positive test result with the target sequence for which the module was developed. This should be demonstrated by *in silico* similarity searches against available and appropriate databases (e.g. EMBL, GenBank, DDBJ, Patent, etc.) and experimentally with suitable samples with and without the presence of the target sequence.

Note: The definition is adopted from EURL-GMFF (2008).

Note: the definition of appropriate database is sector dependent. The target organism(s) is evidently always to be included. For GMO plants the databases should typically include all available sequences of transgenic constructs, agriculturally important plant species, soil, water and plant associated viruses, microorganisms and invertebrate animal species. For food pathogens the databases should typically include all available sequences of soil, water, plant, animal and food associated viruses, microorganisms and invertebrate animal species, as well as domesticated vertebrate animals. For the detection of plant and animal species the database should typically include all available related allelic sequences of plants and animals, as well as all other available sequences of plants, animals, viruses and microorganisms.

AAV: *In silico*, it must be demonstrated that the target sequence is unique, and any similarities that theoretically can result in false positives with the analytical module must be reported, and its potential consequences must be briefly discussed.

Note: For a PCR module false positives are theoretically possible mainly if the number of intended mismatches in one or more of the oligonucleotides (primers, probes) is low.

Note: For a sequencing module the known and expected differences between the target(s) and non-targets must be described.

AAV: Experimentally, at the concentrations specified below, it must be demonstrated that no false positives are produced with the samples that do not contain the target sequence, and that in phase *i* (development) at least 95% and in phase *ii* (validation) all samples containing the target sequence yields a positive test result.

Note: These tests should be conducted with approximately 4,000 copies or 10 ng of non-target DNA and with between 100 and 1000 copies or between 0.25 and 2.5 ng of target DNA, unless a more relaxed or strict AAV is justified (e.g. by reference to contractual or legal requirements).

Note: These concentrations are high enough to give a realistic impression of the exclusivity and inclusivity of the module.

Note: The mass indicated above is corresponding to a genome size of approx. 3×10^9 bp (1C mass approx. 3 pg, i.e. genome sizes similar to maize and human). For significantly smaller or larger genome sizes the mass should be adjusted according to (estimated) genome size.

Note: Sometimes the inclusion of very high quantities of non-target DNA is recommended to exclude cryptic cross-reactivity. Plant pathologists for example test specificity against 10^6 CFU of non-target bacteria.

Note: A description must be included in the module, explaining how a true positive can be discriminated from a false positive. Appropriate techniques ranked with respect to expected reliability are: 1) DNA sequencing; 2) probe hybridisation; 3) melting temperature profiling or restriction enzyme digestion and fragment size analysis.

Note: The terms inclusivity and exclusivity are sometimes used, e.g. in microbiology. Inclusivity is the ability to detect the target and exclusivity is the ability to discriminate against (by not detecting) non-targets.

Note: For food and feed microbiology at least 30 diverse target strains and at least 20 non-target strains of diverse but close relatives to the target shall be included in the test.

AAV: For taxon-specific modules the absence or degree of allelic variation across a globally representative collection of the taxon should be demonstrated experimentally.

Note: For GMO plants, a globally representative collection is typically at least 20 phylogenetically distantly related cultivars (lines) of the taxon. For food pathogenic bacteria, a globally representative collection is typically at least 100 distantly related strains of the taxon. For detection of taxa for customs and excise duty, a globally representative collection should consist of a statistically representative sample of breeds or subspecies among those typically entered into commerce. Any deviations from these guidelines should be explained and justified (e.g. with reference to regulatory or contractual requirements).

Note: For multiplex qualitative modules with capacity to discretely identify individual targets, specificity should be evaluated for each target sequence, while using the final multiplex setup. The same acceptance criteria as for the corresponding singleplex modules should then be applied.

3.3. False positive rate (type I error rate)

Definition: The probability α of making a type I error (scoring a false positive).

Note: A false positive occurs if the test result is classified positive (target is detected) when the actual condition is negative (target is absent).

Note: In digital PCR two types of false positive results can occur: 1) individual false positive partitions, and 2) a false conclusion about a positive test result for the whole sample. The latter occurs if the rate of (false) positive partitions exceeds the validated false positive rate of the module.

Acceptance criterion (AAV): $\alpha \leq 5\%$ at all concentrations, unless a more relaxed or strict AAV is justified (e.g. by reference to contractual or legal requirements).

Note: $\alpha = 100 \times$ number of misclassified known negative samples/total number of known negative samples. A minimum of 20 diverse samples must be tested representing closely related taxa, representative co-occurring impurities (e.g. viruses, endo- and episymbionts) and strains with molecular make-up challenging the exclusivity of the module (e.g. alleles differing only at a few nucleotide positions).

Note: The concentrations to be tested should at least include the lower and upper concentrations at which the module is intended for use, and two other concentration levels within that range.

Note: Classification should be done applying the final assay layout and evaluation algorithm.

Note: For food pathogen detection it is required that $\alpha < 1\%$ at all concentrations ≥ desired LOD.

Note: In some cases the positive test result is caused by actual presence of the target in the sample. Some certified reference materials (CRMs) for GMO detection is an example. These CRMs are certified for the presence and quantity of a particular item (e.g. GMO A), but not for the presence or absence of other items (e.g. GMO B). Prudence in the interpretation of unexpected positive results is therefore required.

Note: In the case of digital PCR, $\alpha \leq 0.2\%$ for individual partitions and $\alpha \leq 5\%$ for the whole sample.

3.4. False negative rate (type II error rate)

Definition: The probability β of making a type II error (scoring a false negative).

Note: A false negative occurs if the test result is negative (target is not detected) when the actual condition is positive.

Note: In digital PCR two types of false negative results can occur: 1) individual false negative partitions, and 2) a false conclusion about a negative test result for the whole sample. The latter occurs if the rate of (false) negative partitions exceeds the validated false negative rate of the module.

Acceptance criterion (AAV): $\beta \leq 5\%$ at all target concentrations ≥ desired LOD unless a more relaxed or strict AAV is justified (e.g. by reference to contractual or legal requirements).

Note: $\beta = 100 \times$ number of misclassified known positive samples/total number of known positive samples. A minimum of 20 diverse samples must be tested representing the largest available diversity of target containing strains or taxa, to challenge the inclusivity of the module.

Note: The concentrations to be tested should at least include the lower and upper concentrations at which the module is intended for use, and two other concentration levels within that range.

Note: Classification should be done applying the final assay layout and evaluation algorithm.

Note: For food pathogen detection it is required that $\beta < 1\%$ at all concentrations ≥ desired LOD.

Note: In the case of digital PCR, $\beta \leq 0.2\%$ for individual partitions and $\beta \leq 5\%$ for the whole sample.

3.5. Limit of detection (LOD)

Definition: The limit of detection is the lowest amount or concentration of analyte in a sample, which can be reliably detected, but not necessarily quantified.

Acceptance criterion (AAV): The LOD is the lowest concentration at which the ***probability of detection (POD)*** ≥ 95%, i.e. the lowest concentration yielding a ***false negative rate*** ≤ 5%, unless a more relaxed or strict AAV can be justified (e.g. by reference to regulatory or contractual requirements).

Note: The definition is modified from EURL-GMFF (2008) and ISO (2006).

Note: For GMO the AAV is invariant for individual (e.g. a transgene or taxon specific) and combined (transgene + taxon specific) modules.

Note: Theoretically the absolute (copy number based) LOD of a PCR module is 5-10 target copies. The absolute LOD of a PCR module should therefore be ≤ 25 target copies, unless a more relaxed or strict AAV is justified (e.g. by reference to contractual or legal requirements).

For multiplex qualitative PCR modules an additional acceptance criterion can be required:

3.5.1. Asymmetric LOD (LODasym)

Definition: the minimum ratio between the concentration or copy number of the primary target (sequence) and the concentration or copy number of the secondary target (sequence[s]) at which the primary target can still be detected with a ***probability of detection (POD)*** ≥ 95%, i.e. yielding a ***false negative rate*** ≤ 5%.

Acceptance criterion: The LODasym shall be < 1:1000, unless a more relaxed or strict AAV can be justified (e.g. by reference to regulatory or contractual requirements).

Note: The LODasym is determined by testing the primary target at low concentration (corresponding or close to the absolute [copy number based] LOD) in the presence of increasing concentrations of the secondary target(s) in the multiplex assay. For multiplex assays with many targets this (checkerboard) approach can, however, lead to unfeasibly large numbers of combinations to be tested. In such cases more pragmatic approaches can be justified. One example approach is to first determine the amplification efficiency and LOD for each individual target, and then verify that the LOD is still acceptable for the least efficiently amplified target in the presence of high concentrations of the most efficiently amplified target.

Note: Molecular methods detect and measure DNA (or RNA) molecules, not mass units, cells, CFUs or viral particles. However, as the latter cannot always be translated into DNA copies, the LODasym can also, if necessary, be expressed in terms of mass, cell or CFU ratios, or equivalent ratios.

Note: as an example, in a duplex PCR module, if 20 copies of the primary target sequence are detected with POD ≥ 95%, in presence of > 20,000 copies of secondary target sequence, the LODasym is then a ratio below 1:1000.

Note: In case of multiplex PCR modules, the LODasym can be determined by testing e.g. 20 copies of each target sequence (serving as a primary target) in presence of a background of all other targets (serving as secondary target) summed at the level of 20,000 copies.

Note: For some qualitative multiplex assays the purpose is only to determine if at least one target is present, without a need for detection of all targets present in the sample. If the positive test result is verified, e.g. by sequencing, then the LODasym for individual primary targets do not have to be established, provided the LOD for each individual target is acceptable.

3.6. Lower limit of quantification (LOQ)

Definition: The lower limit of quantification is the lowest amount or concentration of analyte in a sample that can be reliably quantified with an acceptable level of ***accuracy*** (3.8).

Acceptance criterion: the LOQ should be = the lowest amount or concentration included in the ***dynamic range*** (3.7).

Note: The definition is modified from EURL-GMFF (2008) and ISO (2006).

Note: The LOQ should be assessed experimentally. Estimates of LOQ should be obtained on a sufficient number of test results. For a full validation this should be at least 15, by analogy with the requirement set for the assessment of RSDr. This allows estimating the LOQ in conjunction with the assessment of RSDr.

Note: The upper end of the dynamic range of a quantitative method corresponds to the highest amount or concentration of analyte that can be reliably quantified with an acceptable level of trueness and precision. Although this *de facto* would correspond to an upper limit of quantification, in scientific literature the LOQ is usually only referring to the lower limit of quantification. This approach is also adopted here.

Note: LOQ is not applicable to qualitative methods.

3.7. Dynamic range

Definition: The range of concentrations over which a quantitative module performs in a linear manner with an acceptable level of ***accuracy*** (3.8).

Acceptance criterion: The dynamic range should cover the full concentration range that the module is intended for. This can be expressed as a relative concentration (% relative to a specified unit) or absolute concentration (copy number range).

Note: The definition is modified from EURL-GMFF (2008).

Note: Dynamic range is only applicable to quantitative modules. However, the applicability range, i.e. the range of concentrations at which a qualitative module can be applied is a related parameter which is not defined in the present version of this guidance document.

Note: Relative concentration = e.g. 0.01% - 10% for SNPs or GMOs. Absolute concentration = e.g. 10^2 – 10^7 copies per test (PCR, sequencing reaction, etc.).

Note: For quantitative real-time PCR modules the dynamic range is established on the basis of a standard curve tested on a minimum of 4 (in phase *i*; development), respectively 5 (in phase *ii*, validation) concentration levels evenly distributed over at least the full concentration range that the module is intended for and with at least 3 (in phase *i*; development), respectively 6 (in phase *ii*, validation) repetitions. A more relaxed or strict AAV must be justified (e.g. by reference to contractual or legal requirements).

Note: For digital PCR (see chapter 4, glossary for definition) platforms the maximum input of template DNA per PCR is often far more limited than for quantitative real-time PCR modules. However, some are comparable to quantitative real-time PCR platforms. For digital PCR the linear response is demonstrated by testing of at least 5 concentration levels evenly distributed over at least the full concentration range that the module is intended for and with a sufficiently high number of repetitions (PCRs) to enable the estimation of ***precision*** (3.10). Independently of the number of partitions in the digital PCR, the test result is considered positive if the number of partitions yielding a positive test response (target detected) exceeds the false positive rate. Moreover, when the majority of partitions yield positive test responses, the number of partitions yielding a negative test response (target not detected) must exceed the false negative rate to be able to calculate a target concentration (estimate). A more relaxed or strict AAV must be justified (e.g. by reference to contractual or legal requirements).

3.8. Accuracy (trueness and precision)

Definition: The closeness of agreement between a test result and the accepted reference value (Figure 2). Accuracy is expressed on the basis of combined ***trueness*** and ***precision***.

Acceptance criterion (AAV): Acceptable accuracy is obtained when both ***trueness*** (3.9) and ***precision*** (3.10) meet their respective AAVs.

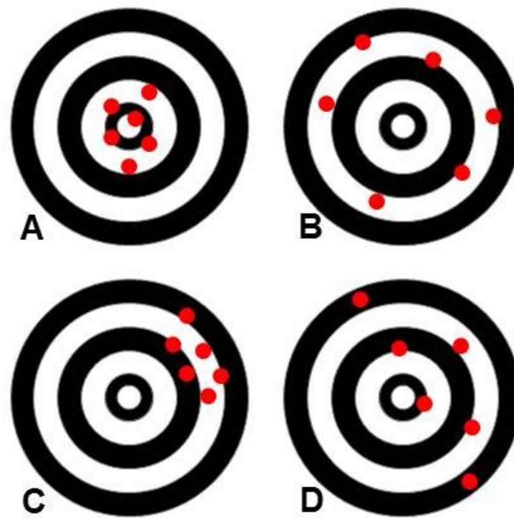


Figure 2. Accuracy, trueness and precision. Accuracy is decreasing from top left to bottom and right. This is due to decreasing trueness (top to bottom) and decreasing precision (left to right).

3.9. Trueness

Definition: The closeness of agreement between the average value obtained from a large series of test results and an accepted reference value. The measure of trueness is usually expressed in terms of bias. Trueness can therefore be described as the accuracy of the mean.

Acceptance criterion (AAV): The trueness should be within ± 35% of the accepted reference value over the whole ***dynamic range*** of the module, unless a more relaxed or strict AAV can be justified (e.g. by reference to regulatory or contractual requirements).

Note: The definition inevitably requires a large series of test results (minimum 5 concentration levels and 6 repetitions for phase *ii*; validation). However, estimates of trueness can be obtained in phase *i* (development) with a smaller number of test results (minimum 4 concentration levels and 3 repetitions).

Note: For GMO, the current AAV according to EU regulation and EU guidance documents is ± 25% for individual (e.g. transgene or taxon specific) as well as combined (e.g. transgene + taxon specific) PCR modules (EURL-GMFF 2008).

Note: The definition is adopted from EURL-GMFF (2008) and ISO (2006).

Note: Trueness is not applicable to qualitative methods.

Note: Trueness is theoretically better at higher concentrations than at near LOQ concentrations. It is possible to define the AAV for trueness as a range which is narrower at high concentrations than at low concentrations.

3.10. Precision

Definition: The closeness of agreement between independent test results obtained under stipulated conditions.

Note: The definition is adopted from AOAC (2012a) and ISO (1994a; 2006).

Note: Representative examples of stipulated conditions are repeatability conditions (used to determine the within laboratory precision) and reproducibility conditions (used to determine the between laboratory precision).

Taxon-specific modules intended for quantitative analysis should target alleles for which the copy number per haploid genome is already determined (preferably single copy). The degree and preferable absence of copy-number variation across a globally representative collection of the taxon should be demonstrated and preferably be reported with a brief discussion of potential consequences.

Note: With quantitative real-time PCR the within-taxon range of variability of Cq values with an invariant mass of DNA from the taxon in question should be:

$Cq_{mean} - Cq_{min} < 1.5$;

$Cq_{max} - Cq_{mean} < 1.5$; and

$Cq_{max} - Cq_{min} \leq 2.0$

unless a more relaxed or strict AAV is justified (e.g. by reference to contractual or legal requirements).

Note: Precision depends only on the distribution of random errors and does not relate to the true value or to the specified value.

Note: The measure of precision usually is expressed in terms of imprecision and computed as a standard deviation of the test results. Less precision is reflected by a larger standard deviation.

Note: Independent test results means results obtained in a manner not influenced by any previous result on the same or similar test object. Quantitative measures of precision depend critically on the stipulated conditions.

Note: Correct estimates of precision are theoretically more reliably obtained at high concentrations than at near LOQ concentrations as well as with high numbers of test results than with low numbers of test result. It is possible to define the AAV for precision as a range which is narrower at high concentrations than at low concentrations.

3.10.1. Relative repeatability standard deviation (RSD_r) – within laboratory precision

Definition: The relative standard deviation of test results obtained under repeatability conditions. Repeatability conditions are conditions where test results are obtained with the same module, on identical test items, in the same laboratory, by the same operator, using the same equipment within short intervals of time.

Acceptance criterion (AAV): The relative repeatability standard deviation should be ≤ 35% over the whole dynamic range, unless a more relaxed or strict AAV can be justified (e.g. by reference to contractual or legal requirements).

Note: The definition is adopted from AOAC (2012) and EURL-GMFF (2008).

Note: For GMO, the current AAV according to EU regulation and EU guidance documents is $RSD_r \leq 25\%$ for individual (e.g. transgene or taxon specific) as well as combined (e.g. transgene + taxon specific) PCR modules (EURL-GMFF 2008).

Note: RSD_r is relevant for phase *i* (development), *ii* (validation) and *iii* (application of module).

Note: Estimates of repeatability should be obtained on a sufficient number of test results applying the final assay layout. For phase *i* (development) the number of test results should be at least 3 at each concentration level. For a full validation (phase *ii*) the number of test results should be at least 15, based on ISO 5725-3 (ISO 1994b).

Note: RSD_r is not applicable to qualitative methods.

3.10.2. Relative reproducibility standard deviation (RSD_R) – inter laboratory precision

Definition: The relative standard deviation of test results obtained under reproducibility conditions. Reproducibility conditions are conditions where test results are obtained with the same module, on identical test items, in different laboratories, with different operators, using different equipment. Reproducibility standard deviation describes the inter-laboratory variation.

Acceptance criterion: The relative reproducibility standard deviation RSD_R should be < 50% over the whole dynamic range. At low concentrations (e.g. < 250 copies or < 0.2% relative concentration) RSD_R values < 67% are deemed acceptable. More relaxed or strict AAV must be justified (e.g. by reference to contractual or legal requirements).

Note: The definition is adopted from AOAC (2012) and EURL-GMFF (2008).

Note: For GMO, the current AAV according to EU regulation and EU guidance documents is $RSD_R \leq 35\%$ over the whole dynamic range for individual (e.g. transgene or taxon specific) as well as combined (e.g. transgene + taxon specific) PCR modules), except at concentrations < 0.2% where $RSD_R \leq 50\%$ is acceptable (EURL-GMFF 2008).

Note: RSD_R is determined in phase *ii* (validation), often in a collaborative trial. The minimum number of test results is 15 per concentration level. RSD_R can be an important parameter for selection of methods for phase *iii* (routine application).

Note: RSD_R is usually calculated only for quantitative modules. However, the ***Probability of Detection (POD***; 3.11) may be determined by interlaboratory studies and can play a similar role in the assessment of module reliability.

3.11. Probability of Detection (POD)

Definition: The proportion of positive results for a qualitative module at a given analyte concentration.

Note: The definition is modified from AOAC (2012).

Note: POD is concentration dependent.

Note: The assessment of interlaboratory probability of detection (POD) can be done in a collaborative study. To obtain data for determination of the POD, different laboratories measure replicate DNA samples at a suitable range of target concentrations. The range of concentrations should at least include the lowest and highest concentrations at which a POD ≥ 95% is desired, unless more relaxed or strict AAVs can be justified (e.g. by reference to contractual or legal requirements).

Note: With an appropriate design of the collaborative study, POD response curves can be calculated to provide an overview of the performance of the module. This response curve should be compared to the “ideal curve” calculated on basis on the underlying probability (Poisson) distribution of the target across the concentration series. The POD can also be calculated for all laboratories or separately for each laboratory to identify outliers.

Note: the relations between the average probability of detection across laboratories, the reproducibility standard deviation of the probability of detection and the interval within which the laboratories probabilities of detection are expected to lie are not the same as the analogous ***relative reproducibility standard deviation (interlaboratory precision;*** RSD_R***; 3.10.2)*** of quantitative methods determined in collaborative trials.

Acceptance criterion: POD ≥ 0.95 (95% probability of detection) for the full range of concentrations at which the module is intended to be applicable, unless a more relaxed or strict AAV can be justified (e.g. by reference to contractual or legal requirements).

Note: POD can be assessed in addition to, but not replacing, ***false positive rate*** (3.3) and ***false negative rate*** (3.4) and provides additional information on the performance of a qualitative module.

4. Glossary:

Digital PCR = a PCR approach using the positive (detected) vs. negative (not detected) qualitative test results from a high number of partitions (parallel PCRs) and Poisson distribution to obtain quantitative estimates of the concentration of target sequence.

> **Note:** Here we define the minimum number of partitions of a digital PCR (dPCR) = 400, to discriminate digital PCR from most probable number PCR (MPN-PCR). MPN-PCR is related to dPCR but uses much lower numbers of partitions, depends on diluting the template nucleic acid to a working concentration of approximately 1 target copy/PCR and a different statistical approach to calculations of concentrations.

Method = a combination of modules that together form a complete set of operations that permit the analytical laboratory to process a received sample, do the analysis and obtain and report a final test result.

Module = a distinct and limited operation that is performed on an input material and which delivers an altered output material or data.

> **Note:** Examples of modules are: 1) a sample preparation module where the input material is e.g. grains and the output material could be flour; 2) a DNA extraction and purification module where the input material is flour and the output material is purified DNA in aqueous solution; 3) a real-time PCR module where the input material is a purified DNA in aqueous solution and the output material is a measurement of fluorescence and translation into a number of target sequence copies; or 4) a data evaluation module where the available data are processed into a final measurement result.
>
> **Note:** Definition adopted from Holst-Jensen et al. (2012).

Practical limit of detection (pLOD) = The pLOD is the lowest relative quantity (concentration) of the primary target sequence that can be detected, given a known (determined/estimated) number of copies of a secondary target sequence.

> **Note:** Definition modified from EURL-GMFF (2012).
>
> **Note:** pLOD is case (sample) specific.

Practical limit of quantification (pLOQ) = The pLOQ is the lowest relative quantity (concentration) of the primary target sequence that can be reliably quantified, given a known (determined/estimated) number of copies of a secondary target sequence.

> **Note:** Definition modified from EURL-GMFF (2012).
>
> **Note:** pLOQ is case (sample) specific.

Primary target sequence = a target sequence representing a smaller group which is a member of a larger group represented by a secondary target sequence.

> **Note:** The following three examples illustrate the difference between a primary and secondary target sequence:
>
> For GMO, a transgene specific marker is a primary target sequence and a species specific marker is a secondary target sequence.
>
> For STEC, a serotype or pathogen marker is a primary target sequence and an Enterobacteriaceae specific marker is a secondary target sequence.
>
> For customs, a plant species specific marker is a primary target sequence and a chloroplast specific marker is a secondary target sequence. Chloroplast specific markers are usually unfit for quantitative analyses.

Secondary target sequence = a target sequence representing a larger group which includes but is not limited to a group represented by a primary target sequence.

Note: See examples given under primary target sequence

Target sequence = the sequence motif or group of sequence motifs that the module is designed to detect

Note: For example a specific DNA sequence motif defined by a set of terminal primer motifs and an internal probe motif.

Test = analytical experiment performed with the purpose to determine whether the corresponding target sequence is present or absent in the material subject to analysis.

Note: The test is performed with a specific testing module, e.g. a PCR module with a specified set of primers and probe, reagent concentrations and cycling profile.

Note: Definition adopted from EURL-GMFF (2012).

5. References

5.1. Papers cited in the report

AOAC International (2012). Official Methods of Analysis (OMA), appendix f. http://www.eoma.aoac.org/appendices.asp.

EURL-GMFF (2008). Definition of minimum performance requirements for analytical methods of GMO testing. Developed by the European Network of GMO Laboratories (ENGL). http://gmo-crl.jrc.ec.europa.eu/doc/Min_Perf_Requirements_Analytical_methods.pdf (accessed 20140818).

EURL-GMFF (2012). Overview on the detection, interpretation and reporting on the presence of unauthorised genetically modified materials. Developed by the European Network of GMO Laboratories (ENGL). http://gmo-crl.jrc.ec.europa.eu/doc/2011-12-12%20ENGL%20UGM%20WG%20Publication.pdf (accessed 20140818).

European Commission (2003). Regulation (EC) No 1829/2003 of the European Parliament and of the Council of 22 September 2003 on genetically modified food and feed. Official Journal of the European Union L 268: 1-23.

Holst-Jensen A, Bertheau Y, de Loose M, Grohmann L, Hamels S, Hougs L, Morisset D, Pecoraro S, Pla M, Van den Bulcke M, Wulff D. (2012**)**. Detecting un-authorized genetically modified organisms (GMOs) and derived materials. Biotechnology Advances 30 (6): 1318-1335 doi: 10.1016/j.biotechadv.2012.01.024

ISO (1994a). ISO 5725-1. Accuracy (trueness and precision) of measurement methods and results -- Part 1: General principles and definitions. International Organization for Standardization. 17 pp.

ISO (1994b). ISO 5725-3. Accuracy (trueness and precision) of measurement methods and results -- Part 3: Intermediate measures of the precision of a standard measurement method. International Organization for Standardization. 25 pp.

ISO (2006). ISO 24276. Foodstuffs -- Methods of analysis for the detection of genetically modified organisms and derived products -- General requirements and definitions. International Organization for Standardization. 16 pp.

5.2. Papers consulted but not cited specifically

AOAC International (1999). Qualitative and Quantitative Microbiology Guidelines for Methods Validation. Journal of the AOAC International 82: 402-415.

Asuragen (2014). Functional DNA Quality Analysis Improves the Accuracy of Next Generation Sequencing from Clinical Specimens. http://asuragen.com/wp-content/uploads/2014/01/Next_Generation_Sequencing_WhitePaper-QFI.pdf. (accessed 20140818).

Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, Mills DA, Caporaso JG. (2013) Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nature Methods 10:57-60. doi: 10.1038/nmeth.2276.

Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, Vandesompele J, Wittwer CT. (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clinical Chemistry 55(4):611-22. doi: 10.1373/clinchem.2008.112797.

Cha RS and Thilly WG. (1993). Specificity, efficiency, and fidelity of PCR. Genome Research 3:S18-S29. http://genome.cshlp.org/content/3/3/S18.long.

Dai M, Thompson RC, Maher C, Contreras-Galindo R, Kaplan MH, Markovitz DM, Omenn G, Meng F. (2010). NGSQC: cross-platform quality analysis pipeline for deep sequencing data. BMC Genomics 11 (suppl. 4): S7. doi: 10.1186/1471-2164-11-S4-S7.

Feldsine P, Abeyta C, Andrews WH. (2002). AOAC INTERNATIONAL Methods Committee Guidelines for Validation of Qualitative and Quantitative Food Microbiological Official Methods of Analysis. Journal of the AOAC International 85: 1187-1200. http://www.aoac.org/Official_Methods/Food_Micro_Validation_Guidelines.pdf

Gargis AS, Kalman L, Berry MW, Bick DP, Dimmock DP, et al. (2012). Assuring the quality of next-generation sequencing in clinical laboratory practice. Nature Biotechnology 30(11): 1033-1036. doi: 10.1038/nbt.2403.

Huggett JF, Foy CA, Benes V, Emslie K, Garson JA, Haynes R, Hellemans J, Kubista M, Mueller RD, Nolan T, Pfaffl MW, Shipley GL, Vandesompele J, Wittwer CT, Bustin SA. (2013). Guidelines for

minimum information for publication of quantitative digital PCR experiments. Clinical Chemistry 59 (6): 892-902. doi: 10.1373/clinchem.2013.206375.

Illumina (2011). Quality Scores for Next-Generation Sequencing. Assessing sequencing accuracy using Phred quality scoring. http://res.illumina.com/documents/products/technotes/technote_q-scores.pdf. (accessed 20140818).

ISO (2003). ISO 16140. Microbiology of food and animal feeding stuffs – Protocol for the validation of alternative microbiological methods. International Organization for Standardization.

ISO (2007). ISO 7218. Microbiology of food and animal feeding stuffs – General requirements and guidance for microbiological examinations. International Organization for Standardization.

Keane T. (2010). ECCB 2010 Next-gen sequencing tutorial. http://www.slideshare.net/thomaskeane/eccb-2010-nextgen-sequencing-tutorial#. (accessed 20140818).

Li B, Zhan X, Wing MK, Anderson P, Kang HM, Abecasis GR. (2013). QPLOT: a quality assessment tool for next generation sequencing data. BioMed Research International 2013: article ID 865181. doi: 10.1155/2013/865181.

Meusnier I, Singer GAC, Landry J-F, Hickey DA, Hebert PDN, Hajibabaei M. (2008). A universal DNA mini-barcode for biodiversity analysis. BMC Genomics 9:214. doi: 10.1186/1471-2164-9-214.

NordVal Validation: Protocol for the validation of alternative microbiological methods (2009). http://www.nmkl.org/NordVal/NordValprotocolmarch2009.pdf.

Patel RK, Mukesh J. (2012). NGS QC toolkit: a toolkit for quality control of next generation sequencing data. PLoS ONE 7 (2): e30619. doi: 10.1371/journal.pone.0030619.

Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y. (2012). A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina Miseq sequences. BMC Genomics 13:341. doi: 10.1186/1471-2164-13-341.

Rehm HL, Bale SJ, Bayrak-Toydemir P, Berg JS, Brown KK, Deignan JL, Friez MJ, Funke BH, Hegde MR, Lyon E. (2013). ACMG clinical laboratory standards for next-generation sequencing. Genetics in Medicine 15(9): 733-747. doi: 10.1038/gim.2013.92.

Thermo Scientific (2011) Quality Control during Library Construction for *de novo* Next-Generation Sequencing platforms. http://www.nanodrop.com/Library/T115-Quality-Control-During-Library-Construction-for-de%20novo-Next-Gen-Sequencing-Platforms.pdf. (accessed 20140818).

Trivedi UH, Cezard T, Bridgett S, Montazam A, Nichols J, Blaxter M, Gharbi K. (2014). Quality control of next-generation sequencing data without a reference. Frontiers in Genetics 2014(5): Article 111. doi: 10.3389/fgene.2014.00111.

Zhang T, Luo Y, Liu K, Pan L, Zhang B, Yu J, Hu S. (2011). BIGpre: a quality assessment package for next-generation sequencing data. Genomics, Proteomics & Bioinformatics 9(6): 238-244. doi: 10.1016/S1672-0229(11)60027-2.