Results of the international ring trial for the genetic identification of crustacean species using cytochrome *c* oxidase I and *16S* rRNA gene segments

CHRISTIAN BRENN
&
UTE SCHRÖDER
2021

Report

International ring trial for the validation of an identification method for crustacean species based on COI and *16S rRNA*

In this international ring trial within the Interreg Atlantic Area project SEATRACES, a method for the identification of commercially relevant crustacean species by Sanger sequencing of fragments of the genetic marker cytochrome-*c*-oxidase I and *16S* rRNA was carried out for validation purposes.
Beforehand, a preliminary test was carried out within the SEATRACES consortium with seven laboratories, which showed positive results so that an international ring trial could be started. A selection of food control laboratories located within the respective countries were suggested by members of the SEATRACES consortium and subsequently contacted by the Max Rubner-Institut. In the end, twelve laboratories from five EU countries (1x Spain, 1x Portugal, 1x Ireland, 1x England, 8x Germany) took part in the ring trial.
The participants received a standard operating procedure (SOP) (see Annex I) beforehand, explaining in detail the procedure for the proposed method. The laboratories were also informed that they needed to provide the necessary reagents themselves, except primers. Recommendations were given in the SOP but laboratories were free to choose their preferred PCR reagents or method of choice for DNA extraction/clean-up. Nine tissue samples á 200-250 mg in EtOH (96%), as well as two pairs of primers (Lobo et al. 2013, Palumbi et al. 1991) were provided by the Max Rubner-Institut, with every laboratory receiving 10 µL (100 pmol/µL) of each primer. The forward and reverse DNA-strands of COI and 16S should be sequenced by the participants before combining them into a consensus sequence. The edited sequences should then be compared with reference sequences in GenBank using the BLAST algorithm. COI sequences should additionally be compared with reference sequences in the BOLD database.
Participants were asked to note any deviations from the SOP in order to assess robustness of the method and to be able to evaluate the results of different labs obtained with different reagents, extraction methods, etc.

Sample species were chosen based on their commercial relevance and taxonomic diversity. Species that are known to be “problematic” during DNA amplification were also included (Table 1)

Table 1. Samples of the international ring trial

1	*Pleoticus muelleri*
2	*Penaeus monodon*
3	*Nephrops norvegicus*
4	*Macrobrachium rosenbergii*
5	*Squilla mantis*
6	*Procamberus clarkii*
7	*Homarus americanus*
8	*Litopenaeus vannamei*
9	*Nephrops norvegicus*

Methods for DNA extraction (CTAB or QIAamp DNA extraction Kit (Qiagen)) and a PCR mastermix (HotStarTaq Mastermix Plus (Qiagen)) were recommended in the SOP. However, the participating laboratories were free to choose their method of choice for DNA extraction or their preferred PCR reagents. Any deviations from the SOP regarding the DNA extraction needed to be noted in the results document (Table 2).

Table 2. Overview of the DNA extraction methods used by the participating laboratories.

Lab	DNA extraction method/ kit	Comments
1	Wizard Kit (Promega)	
2	DNeasy Blood&Tissue Kit (Qiagen)	
3	DNeasy maricon Food Kit (Qiagen)	
4	CTAB extraction	
5	QIAamp DNA Mini Kit (Qiagen)	
6	NucleoMag Vet Kit (Macherey-Nagel)	
7	Maxwell RSC Purefood GMA and Authentication Kit (Promega)	
8	CTAB extraction	Followed by NucleoMag Plant Kit (Macherey & Nagel)
9	NucleoSpin Food Kit (Macherey & Nagel)	
10	QIAamp DNA Mini Kit (Qiagen)	
11	NucleoSpin Food Kit (Macherey & Nagel)	
12	QIAamp DNA Mini Kit (Qiagen)	

Any deviations from the SOP regarding the PCR reagents were to be noted in the results document (Table 3).

Table 3. Overview of the PCR mastermixes used by the participating laboratories.

Lab	Mastermix	Comments
1	BIOTAQ DNA Polymerase (Bioline)	
2	Platinum II Hot-Start PCR Mastermix (ThermoFisher)	
3	HotStarTaq Mastermix (Qiagen)	No deviations were stated
4	Ex Taq DNA Polymerase Hot-Start (TaKaRa)	
5	No Rox Probe Core Kit dTTP (Takyon)	
6	HotStarTaq Mastermix (Qiagen)	No deviations were stated
7	HotStarTaq Mastermix (Qiagen)	No deviations were stated
8	HotStarTaq Mastermix (Qiagen)	
9	SYBR Green Super Mix (PerfeCTa)	
10	HotStarTaq Mastermix (Qiagen)	
11	HotStarTaq Mastermix (Qiagen)	No deviations were stated
12	HotStarTaq Mastermix (Qiagen)	No deviations were stated

All laboratories stated deviations from the SOP in some form regarding either the chosen mastermix, the chosen method for DNA extraction or both. Only five laboratories used the recommended methods for DNA extraction with one carrying out a further clean-up step. Regarding the mastermix, eight laboratories did not state any deviations from the SOP or explicitly stated the use of the recommended mastermix. Four laboratories used a different mastermix.
No other major deviations from the SOP were stated by the participants.

Results and discussion

The international ring trial for crustacean species identification by sequencing COI and 16S fragments showed positive results for most of the nine samples (Table 4). With the exception of certain samples which were only identified to the genus level by some laboratories, all other samples that were successfully amplified and sequenced were also correctly identified to the species level.

Eight of the participating laboratories successfully sequenced all samples with the COI primers, while in four laboratories one or two samples could not be sequenced or it was not possible to obtain amplicons. Samples that were successfully sequenced were mostly correctly identified, with the exception of a few samples that were only identified to the genus level.
Most samples could be successfully identified by the 16S fragment, but only two laboratories were able to correctly sequence and identify all samples. Two laboratories could determine one of the samples (*Macrobrachium rosenbergii*) only to the genus level, while one laboratory stated a second species (*Macrobrachium Idella*) for this sample. In seven laboratories no sequences could be obtained for the *M. rosenbergii*.
In general, no real differences were observable in the frequency of unsuccessful DNA-amplification and sequencing between 16S and COI. Furthermore, no certain specimens or species could not be identified with either of the genetic markers, with the aforementioned exception of *Macrobrachium rosenbergii* with 16S.

Regarding the sequence analysis and editing, it seems that there were some issues and difficulties for some participants. One laboratory did not cut off the primer sequences, in two other laboratories either the 5´-end or 3'-end of the sequences were noticeable shorter after editing than expected. One laboratory seemed to have general problems with sequence editing. In general, it was noticeable that the sequences of the samples differed slightly in length between the laboratories.
It was reported by one laboratory that the identification of three samples with COI and one sample with 16S was done only with the reverse DNA strand, since the sequencing of the forward strand failed or sequences with low quality were obtained.

Table 4. Overview of the international ring trial results

Number of laboratories		12
Number of samples per lab		9
COI GenBank	Number of expected results	108
	Samples with correctly identified species	96
	Samples correctly identified to genus level	4*/**
	Samples without identification	8***
COI BOLD	Number of expected results	108
	Samples with correctly identified species	97
	Samples correctly identified to genus level	4*/**
	Samples without identification	7****/*****
***16S* GenBank**	Number of expected results	108
	Samples with correctly identified species	96
	Samples correctly identified to genus level	3**
	Samples without identification	9

*****All of the obtained sequences allowed species identification.
******One lab assigned one sample to two different species, therefore only "identified to genus".
*******Three of the samples had too low % Identity for reliable species identification, caused by low sequence quality. Laboratories were still able to assign the correct species, but this is not acceptable as species assignment at low identities does not comply with the SOP.
********One sample had too low % Identity for reliable species identification.
*********One sample could not be identified with BLAST or BOLD even if the sequence is identical to the sequences of other laboratories, except for one single nucleotide.

Table 5. Results of COI sequencing.

Species	**Samples correctly identified to species**	**Samples correctly identified to genus**	**Samples without identified species**
Pleoticus muelleri	*9* / 10***	*1****	*2* / 1***
Penaeus monodon	*12* / 11***	*0*	*0* / 1***
Nephrops norvegicus	*10* / 11***	*0*	*2* / 1***
Macrobrachium rosenbergii	*9*	*3****	*0*
Squilla mantis	*9*	*0*	*3*
Procamberus clarkii	*11*	*0*	*1*
Homarus americanus	*12*	*0*	*0*
Litopenaeus vannamei	*12*	*0*	*0*
Nephrops norvegicus	*12*	*0*	*0*

*****Identification with GenBank.
******Identification with BOLD.
*******Species assignment was possible with obtained sequences.

Except for one sample, results of sequence comparison with GenBank were consistent with those of BOLD.
Further analysis of the four samples, which were only identified down to the genus level, showed that an identification down to the species level was possible with comparison of the sequences obtained by the labs and reference sequences in GenBank or BOLD. It is therefore likely that no further detailed investigation of the reference sequences was carried out by the participants.

In one laboratory, the max % identity for *P*. *muelleri* and *N. norvegicus* was quite low with 95% and 96%, respectively, when identifying samples with GenBank. Another laboratory reported a low % identity for *S. mantis* which was probably caused by the low quality of the sequence. Both laboratories were still able to assign the correct species but this is not acceptable as species assignment at low identities does not comply with the SOP

One laboratory reported that only the reverse strand was used for identification of certain samples: *H. americanus*, *N*. *norvegicus* (Sample 9)
Three labs reported low level background sequence traces or several faint bands for some samples, when PCR amplicons were analysed in an agarose gel. The latter did not seem to be a problem for sequencing or analysis of obtained sequences.

Table 6: Results of *16S* sequencing.

Species	**Samples correctly identified to species**	**Samples correctly identified to genus**	**Samples without identified species**
Pleoticus muelleri	*11*	*0*	*1*
Penaeus monodon	*12*	*0*	*0*
Nephrops norvegicus	*12*	*0*	*0*
Macrobrachium rosenbergii	*2*	*2*	*8*
Squilla mantis	*12*	*0*	*0*
Procamberus clarkii	*12*	*0*	*0*
Homarus americanus	*11*	*1**	*0*
Litopenaeus vannamei	*12*	*0*	*0*
Nephrops norvegicus	*12*	*0*	*0*

*****Species assignment was possible by checking the neighbor-joining tree view of aligned/blasted sequences.

Besides the *M*. *rosenbergii* sample, almost all laboratories were able to obtain sequences for all other samples and to assign the correct species. One laboratory only identified the *H*. *americanus* sample up to the genus level.
M. rosenbergii seemed to be problematic as template for the 16S primer. Only five labs were able to obtain sequences from which two were identified up to the genus level and just two were correctly identified.

Notes regarding the examined species:

1. *Pleoticus muelleri*

- Four COI sequences of *Pleoticus robustus* with questionable species assignment were found in Genbank (e.g. KJ879318.1) and in BOLD (Mined from Genbank). These *P. robustus* sequences were probably assigned to the wrong species, since they are all from the same author and cannot be aligned with other valid *P*. *robustus* sequences but with other valid *P*. *muelleri* sequences. When blasting sequences of *P*. *muelleri* it is very likely to find those questionable "*P*. *robustus"* sequences with >98% identity. This problem seems to have been recognised by most laboratories.
- One laboratory could identify *P. muelleri* with only 95% identity when using GenBank/BLAST. Identification with >98% identity was possible using BOLD.

2. *Penaeus monodon*

- One laboratory noticed the occurrence of many *Penaeus* sp. or Penaeide sequences in Genbank with >98% identity. However, this is not considered problematic since enough valid reference sequences of *P*. *monodon* can be found in the databases
- As it was reported before, it is noticeable that that two variants of sequences can be found for *P*. *monodon* and therefore a lot of sequences <98% when using GenBank and BLAST (Brenn et al., 2021). This can be explained by the fact that these sequences are either genetically very different haplotypes of *P*. *monodon* or, as is also suspected, a cryptic species, which is unfortunately also listed as *P*. *monodon* (Yudhistira & Arisuryanti, 2019). It is currently not possible to say exactly which of the two sequence variants is *P*. *monodon* and which is the cryptic species, if it is one at all.
- One lab could not identify the *P. monodon* samples with BOLD. The sequence was almost identical to the sequences of other participants except for one nucleotide. Further investigations showed that after removal of the one nucleotide by editing, it was possible to identify the sample again. Therefore, it is assumed that this problem is caused by the BOLD search algorithm.

3. *Nephrops norvegicus*

- Two samples of *N. norvegicus* were chosen for this international ring trial. It was noticeable that one laboratory had problems generating good quality COI forward-sequences for one of the two samples (Sample 9), while a different laboratory had problems with the other sample (Sample 3). No such problems occurred with the 16S primers.
- One laboratory could identify *N. norvegicus* with only 96% identity when using GenBank/BLAST. Identification with >98% identity was possible using BOLD.

4. *Macrobrachium rosenbergii*

- *M. daqueti* is a synonym for the accepted name *M. rosenbergii* (World Register of Marine Species database)
- Most laboratories assigned the correct species to this sample based on the COI-sequence (2x identification to genus level, 1x two species stated (*M. rosenbergii* / *M. rude*) compared to the 16S sequence. The 16S primers do not seem to be suitable for *M*. *rosenbergii* since only two laboratories could successfully generate sequences and assign the correct species. One laboratory could identify the sample to the genus level and one lab stated two species (*M. rosenbergii* / *M*. *idella*), while the remaining eight laboratories could not generate 16S PCR amplicons or sequences with a quality for reliable identification.
- One laboratory identified the sample with COI additionally as *M. rude*. Since only one COI sequence of *M*. *rude* can be found in GenBank (99,84%, MK792418), but multiple ones of *M*. *rosenbergii* it can be assumed that the *M*. *rude* sequence was probably assigned to the wrong species. Still, there is the possibility that species differentiation might not be possible if valid *M*. *rude* sequences >98% are uploaded to GenBank.

- One laboratory identified the sample with 16S additionally as *M*. *idella*, while other laboratories identified the sample only to genus level since one *M. Idella* sequence > 98% ID can be found in GenBank. It can be assumed that the one *M*. *idella* sequence has an incorrect species assignment since it can be aligned with *M*. *rosenbergii* sequences but not with other *M*. *idella* sequences < 98% ID.

5. *Squilla mantis*

- Two laboratories had no results for COI, while another laboratory assigned the correct species but with only 85% identity due to a lot of ambiguous nucleotides within the sequence.
- One laboratory could only use the reverse sequence of 16S, due to low quality of the forward sequence.
- Only three valid 16S sequences for *S*. *mantis* can be found in GenBank.

6. *Procamberus clarkii*

- One laboratory had no result for COI.

7. *Homarus americanus*

- One laboratory could only use the reverse sequence of COI, due to low quality of the forward sequence.
- One laboratory could identify this sample only to genus level with 16S.
- All other laboratories assigned the correct species with 16S. Only two laboratories stated the validation of the assignment by checking the neighbor-joining tree view of the aligned sequences. This is important since multiple *H*. *Gammarus* sequences with >98% identity can be found in GenBank.

8. *Litopenaeus vannamei*

- Two laboratories noted that one sequence of *Metapenaeus* sp. can be found with >98% identity. This sequence cannot be aligned with other *Metapenaeus* spp. sequences and has therefore probably a faulty species assignment.

9. *Nephrops norvegicus*

- see comments above

Conclusions

Overall, the international ring trial of identification of commercially relevant crustaceans based on COI-and 16S fragments can be seen as successful. Eight laboratories successfully obtained COI sequences for all samples and 96 out of 108 samples were correctly identified on species and four samples on genus level
Here, the importance of carefully working with DNA sequences from public databases was once again apparent, as it would have been possible to identify the species of four samples that were only identified down to the genus level, if participants had looked at the reference sequences in more detail or looked at the neighbor-joining tree view of the aligned sequences.

Again, 96 out of 108 samples could be correctly identified with the 16S fragment, which was solely caused by the problem of sequencing *M*. *rosenbergii*. The laboratories that were successful in sequencing even these "problematic" samples used mostly the Wizard DNA Clean up system (Promega) or the QIAamp Mini DNA Kit (Qiagen) for DNA extraction. Apart from this, the results based on 16S were overall quite positive.

In general, the results show that both gene segments are suitable for reliable species identification of crustaceans, while each having their own advantages and disadvantages. It appears that the COI gene segment is suitable for a clearer and more reliable species differentiation of closely related species. When using 16S, no differentiation is possible for certain closely related species or a neighbor-joining tree must be used for verification of results (*H*. *americanus* / *P*. *clarkii*). Use of the COI primers proved to be difficult in some cases, as some laboratories had problems producing amplicons or could only produce low quality sequences.
It has been shown that species which cannot be amplified using the COI primers can be amplified and identified with 16S, while samples that can only be identified to the genus level with 16S can be identified to the species level with COI. It is recommended species identification of "problematic" or completely unknown samples should be performed with both genetic markers, as they complement each other well and allow reliable identification of a broad range of commercially relevant crustacean species.

References

Lobo, J., Costa, P. M., Teixeira, M. A., Ferreira, M. S., Costa, M. H., & Costa, F. O. (2013). Enhanced primers for amplification of DNA barcodes from a broad range of marine metazoans. *BMC ecology*, *13*(1), 34.

Palumbi, S. (1991). Simple fool's guide to PCR.

Brenn, C., Schröder, U., Hanel, R. & Martinez-Arbizu, P. (2021). A multiplex real-time PCR screening assay for routine species identification of four commercially relevant crustaceans. Food Control, 125.

Yudhistira, A., & Arisuryanti, D. T. (2019). Preliminary findings of cryptic diversity of the giant tiger shrimp (Penaeus monodon Fabricius, 1798) in Indonesia inferred from COI mitochondrial DNA. Genetika, 51(1), 251–260.