



Jul 19, 2024

TKI Diatom NGS Protocol

DOI

<https://dx.doi.org/10.17504/protocols.io.6qpvr8k4plmk/v1>

Kevin Beentjes¹

¹Naturalis Biodiversity Center



Kevin Beentjes

Naturalis Biodiversity Center

Create & collaborate more with a free account

Edit and publish protocols, collaborate in communities, share insights through comments, and track progress with run records.

Create free account

OPEN  ACCESS



DOI: <https://dx.doi.org/10.17504/protocols.io.6qpvr8k4plmk/v1>

Protocol Citation: Kevin Beentjes 2024. TKI Diatom NGS Protocol. **protocols.io**

<https://dx.doi.org/10.17504/protocols.io.6qpvr8k4plmk/v1>

License: This is an open access protocol distributed under the terms of the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Protocol status: Working

We use this protocol and it's working

Created: June 19, 2024



Last Modified: July 19, 2024

Protocol Integer ID: 102082

Keywords: dna extraction, pcr, metabarcoding, novaseq, biomonitoring, diatoms, tki diatom ngs protocol, tki diatom ngs protocol this protocol, dna diatom biosensor project, diatom sample, metabarcoding of the rbcl gene, rbcl gene, tki watertechnology, protocol, dna

Abstract

This protocol describes a standardized workflow for the analysis of diatom samples using metabarcoding of the *rbcl* gene. It was created for the DNA Diatom Biosensor project funded by TKI Watertechnologie. More information on the project can be [found here](#).

Field Sampling

- 1 At the selected location, select approximately 5 to 10 reed stems, and cut them at a depth of 15 to 20 cm below the water surface. If there are no reeds present, other (water) plants can also be collected or hard substrate can be scraped off, provided it is in contact with the open water.

Standardized diatom sampling protocols for the Netherlands are recorded in:

- Handboek Hydrobiologie, Boek II (<https://www.stowa.nl/publicaties/handboek-hydrobiologie>)
- NEN 9409: Analyse van Kiezelwieren (<https://www.nen.nl/nen-9409-2015-nl-206039>)

Collect samples for DNA analysis in 96% ethanol, in individually labeled jars or 50ml falcon tubes.

Optional: Collect samples for morphological analysis according to standard protocols.

Diatom Sampling and DNA Extraction

- 2 DNA extraction is done using the DNeasy PowerSoil Pro Kit (Qiagen catalogue number 47016).
For the **ethanol sampling**, containing diatoms that are easily released from the substrate:
 - Shake the jar containing the reed stems vigorously for 30 seconds to homogenize the sample.
 - Remove 2 mL of the mixture into a 2.0 mL eppendorf tube.
 - Spin down the tube briefly (1 minute at maximum speed) to collect the diatoms in a pellet at the bottom of the tube.
 - Pour off the ethanol, air dry briefly and resuspend the pellet in 800 µL of Solution CD1 from the PowerSoil Pro Kit.
- 3 For the **substrate sampling**, containing the diatoms that are not easily released from the substrate in the previous step:
 - Scrape 3 to 5 reed stems along the full length with a scalpel.
 - Add scraped material to the tube with Solution CD1 and the pellet from step 2, or (in case of separate analysis) add the material to a new tube containing 800 µL of Solution CD1 from the PowerSoil Pro Kit.

In some cases, this step may not be possible due to the lack of solid substrate.
- 4 Continue DNA extraction based on manufacturer's protocol included in the kit.
 - Use 100 µL of Solution C6 for the final elution step.

Amplification and Sequencing

- 5 Amplification for metabarcoding on Illumina NovaSeq is performed using a two-step PCR.

The first PCR is used to amplify the rbcL marker region that matches the reference sequences captured in the online Diat.barcode database

(<https://entrepot.recherche.data.gouv.fr/dataset.xhtml?persistentId=doi:10.15454/TOMBYZ>) (Rimet et al., 2019).

Primers rbcL_708F (Stoof-Leichsenring et al., 2012) & rbcL_R3 (Bruder and Medlin, 2007) are used with IDT10 adapters for Illumina sequencing.

- IDT_rbcL_708F:
CACTCTTTCCCTACACGACGCTCTTCCGATCT**AGGTGAAGTTAAAGGTTCATACTTDA**A
- IDT_rbcL_R3:
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**CCTTCTAATTTACCAACA**ACTG

PCR mix (20 µL reaction):

- 10.4 µL mQ water
- 4.0 µL Thermo Fisher Phire Green 5x PCR buffer
- 0.8 µL BSA (10 mg/mL)
- 1.0 µL primer IDT_rbcL_708F (10 µM)
- 1.0 µL primer IDT_rbcL_R3 (10 µM)
- 0.4 µL dNTPs (2.5 mM)
- 0.4 µL Thermo Fisher Phire II HS Taq
- 2.0 µL template DNA (undiluted)

PCR program:

- Initial denature: 30s at 98°C
- Followed by 30 cycles of 5s at 98°C, 5s at 50°C, and 15s at 72°C
- Followed by final extension of 5 minutes at 72°C

- 6 Check the success of the PCR on a 2% agarose gel, for instance the E-Gel 96 Agarose Gels (Thermo Scientific catalogue number G720802).

After the first PCR, clean the products before moving forward to labeling them in the second PCR with a 0.9 ratio of magnetic beads (e.g. Nucleomag NGS clean-up and size select, Macherey-Nagel catalogue number 744970.50). This can be done automatically (for instance with a C.WASH) or manually, by pipetting on a magnetic separation rack or using a magnetic extractor stamp (e.g. a VP 407AM-N1).

Using a magnetic separation rack, the protocol is as follows:

- Let the beads get to room temperature.
- Prepare fresh 80% ethanol (200 µL per sample).

- Mix the beads well by vortexing.
- Add 0.9 x PCR product volume of beads to the wells of a clean 96-well plate.
- Add the PCR product to the MN beads and mix well by pipetting.
- Incubate at room temperature for 5 minutes.
- Place the plate on the magnetic separation rack for 5 minutes.
- When the solution is clear, carefully remove the supernatant. Leave the plate on the magnet.
- Wash the beads twice with 100 μ L 80% ethanol. Leave the plate on the magnet.
- Let the beads air dry for 1 minute. Leave the plate on the magnet.
- Take the plate off the magnet and resuspend the beads with 20 μ L mQ water.
- Place the plate back on the magnetic separation rack for 5 minutes.
- When the solution is clear, move 18 μ L of the the supernatant to a clean plate.

7 The second PCR is used to attach IDT10 indices for Illumina NovaSeq sequencing.

- IDT_i5:
AATGATACGGCGACCACCGAGATCTACAC[i5]ACACTCTTTCCCTACACGACGCTCTTCC
GATC*T
- IDT_i7:
CAAGCAGAAGACGGCATACGAGAT[i7]GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC
*T

PCR mix (20 μ L reaction):

- 11.2 μ L mQ water
- 4.0 μ L Thermo Fisher Phire Green 5x PCR buffer
- 1.0 μ L forward primer (10 μ M)
- 1.0 μ L reverse primer (10 μ M)
- 0.4 μ L dNTPs (2.5 mM)
- 0.4 μ L Thermo Fisher Phire II HS Taq
- 2.0 μ L cleaned template DNA from step 6 (undiluted)

PCR program:

- Initial denature: 30s at 98°C
- Followed by 5 cycles of 5s at 98°C, 5s at 55°C, and 15s at 72°C
- Followed by final extension of 5 minutes at 72°C

8 Measure the amplicon concentrations, for instance on the Agilent Fragment Analyzer, using the NF-910 dsDNA Reagent Kit for 35-1500 bp fragments (according to manufacturer's protocols), followed by a smear analysis using the PROSize 3.0 software on the range between 350 bp to 550 bp.

Using the measured concentrations, now pool each plate of samples equimolarly into a 1.5 mL lowbind eppendorf tube, either manually or using a liquid handling station such as the Qiagen QIAgility or Opentrons OT-2.

Depending on the accuracy of the platform, it is advisable to make sure the lowest and highest concentrations do not differ by more than a factor five. Either (1) pool low and high concentrations first separately, (2) first dilute samples with very high concentration, or (3) process any samples with very low concentrations as if their concentration is 1/5th of the highest concentration. Also include controls (blanks) in the pool.

- 9 The resulting end pool can be cleaned with a 0.9 ratio of magnetic beads (e.g. Nucleomag NGS clean-up and size select, Macherey-Nagel catalogue number 744970.50).

Using a magnetic separation rack, the protocol is as follows:

- Let the beads get to room temperature.
- Prepare fresh 80% ethanol (200 µL per sample).
- Mix the beads well by vortexing.
- Measure the volume of the pool created in the previous step. Always save a portion of the non-cleaned pool (e.g. 100µl) in case something goes wrong during the clean-up.
- Add 0.9 x the used pool volume MN beads to the pool and vortex.
- Incubate at room temperature for 5 minutes.
- Place the tube on the magnetic separation rack for 5 minutes.
- When the solution is clear, carefully remove the supernatant. Leave the tube on the magnet.
- Wash the beads twice with 500 µL 80% ethanol. Leave the tube on the magnet.
- Let the beads air dry for 1 minute. Leave the tube on the magnet.
- Take the tube off the magnet and resuspend the beads with 50-200 µL mQ water.
- Place the tube back on the magnetic separation rack for 5 minutes.
- When the solution is clear, move the the supernatant to a clean 1.5 mL lowbind eppendorf tube.

- 10 When processing multiple plates of samples, each plate will now have its own equimolar pool in a 1.5 mL lowbind eppendorf tube. Measure the concentration of each pool, for instance on the Agilent TapeStation, using the High Sensitivity D5000 ScreenTape Assay (according to manufacturer's protocols). The pools can then be combined equimolarly to a single end-pool.

Sequencing is performed on the Illumina NovaSeq platform in a PE250 run.

Bio-informatic Processing

- 11 Process the demultiplexed NovaSeq sequencing data into ESVs using the APSCALE pipeline (Buchner et al., 2022 / <https://github.com/DominikBuchner/apscale>), using the following settings:
- PE merging: maxdiffpct = 25, maxdiffs = 100, minovlen = 30

- Primer trimming: p5 primer = AGGTGAAGTTAAAGGTTTCATACTTDAA, p7 primer = CCTTCTAATTTACCAACAACACTG, anchoring = false
- Quality filtering: maxEE = 1, min length = 262, max length = 270
- Dereplication pooling: min size to pool = 8
- Denoising: alpha = 2, minsize = 8, to excel = true
- Lulu filtering: minimum similarity = 84, minimum relative cooccurrence = 95, minimum ratio = 1, to excel = true

- 12 To identify the ESVs, perform a BLAST search (Camacho et al., 2009) against the public reference sequences in the Diat.barcode database (<https://entrepot.recherche.data.gouv.fr/dataset.xhtml?persistentId=doi:10.15454/TOMBYZ>). The BLAST tool can be run in the Galaxy environment using a custom script as in Beentjes et al., 2019 (<https://github.com/naturalis/galaxy-tool-BLAST>). To identify ESVs without species-level matches at a higher taxonomic level, combine this BLAST search with a lowest common ancestor analysis, also described in Beentjes et al., 2019 (<https://github.com/naturalis/galaxy-tool-lca>).

The following settings are recommended:

- BLAST: query coverage cutoff = 85, identify % cutoff = 85, maximum number of hits = 100
- LCA: top bitscore % cutoff = 5, minimum bitscore = 170, minimum identity = 85, minimum coverage = 100, settings mode = output the top hit as species identification if it is above the chosen threshold, identity threshold = 98, query coverage threshold = 100

Protocol references

Beentjes, K. K., Speksnijder, A. G., Schilthuizen, M., Hoogeveen, M., Pastoor, R., & van der Hoorn, B. B. (2019). Increased performance of DNA metabarcoding of macroinvertebrates by taxonomic sorting. *PloS one*, 14(12), e0226527. <https://doi.org/10.1371/journal.pone.0226527>

Bruder, K., & Medlin, L. K. (2007). Molecular assessment of phylogenetic relationships in selected species/genera in the naviculoid diatoms (Bacillariophyta). I. The genus *Placoneis*. *Nova Hedwigia*, 85(3–4), 331–352. <https://doi.org/10.1127/0029-5035/2007/0085-0331>

Buchner, D., Macher, T. H., & Leese, F. (2022). APSCALE: advanced pipeline for simple yet comprehensive analyses of DNA metabarcoding data. *Bioinformatics*, 38(20), 4817–4819. <https://doi.org/10.1093/bioinformatics/btac588>

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC bioinformatics*, 10, 1–9. <https://doi.org/10.1186/1471-2105-10-421>

Rimet, F., Gusev, E., Kahlert, M., Kelly, M. G., Kulikovskiy, M., Maltsev, Y., Mann, D. G., Pfannkuchen, M., Trobajo, R., Vasselon, V., Zimmermann, J., & Bouchez, A. (2019). Diat.barcode, an open-access curated barcode library for diatoms. *Scientific Reports*, 9(1). <https://doi.org/10.1038/s41598-019-51500-6>

Stoof-Leichsenring, K. R., Epp, L. S., Trauth, M. H., & Tiedemann, R. (2012). Hidden diversity in diatoms of Kenyan Lake Naivasha: A genetic approach detects temporal variation. *Molecular Ecology*, 21(8), 1918–1930. <https://doi.org/10.1111/j.1365-294X.2011.05412.x>