Project: BONUS BLUEPRINT

Deliverable: 6.1

Title: Report on BCC development

Work package: 6

Deadline: Month 36

Nature: report/protocol

Dissemination level: public

Responsible contact person:

Matthias Labrenz, Leibniz-Institute of Baltic Sea Research Warnemuende, Germany

matthias.labrenz@io-warnemuende.de

Report on BCC development

Introduction/Background

The BONUS BLUEPRINT competence center (BCC) is a virtual operating center where genetic blueprints are generated and analyzed. In this deliverable 6.1, we report on the sampling, laboratory and bioinformatic procedures developed within the BCC. These are prerequisites for acquisition and identification of genetic blueprints and, ultimately, environmental state descriptions, and are depicted in Figure 1. Note, however, that the present report on protocols for sampling, gene sequence acquisition and analysis does not reflect whether genetic blueprints based on taxonomy, specific functional genes, DNA or RNA are best suited as indicators of environmental status. This issue is addressed in parallel work packages, but will be integrated with the final BCC work flow report due in Month 47 of the BONUS BLUEPRINT project.

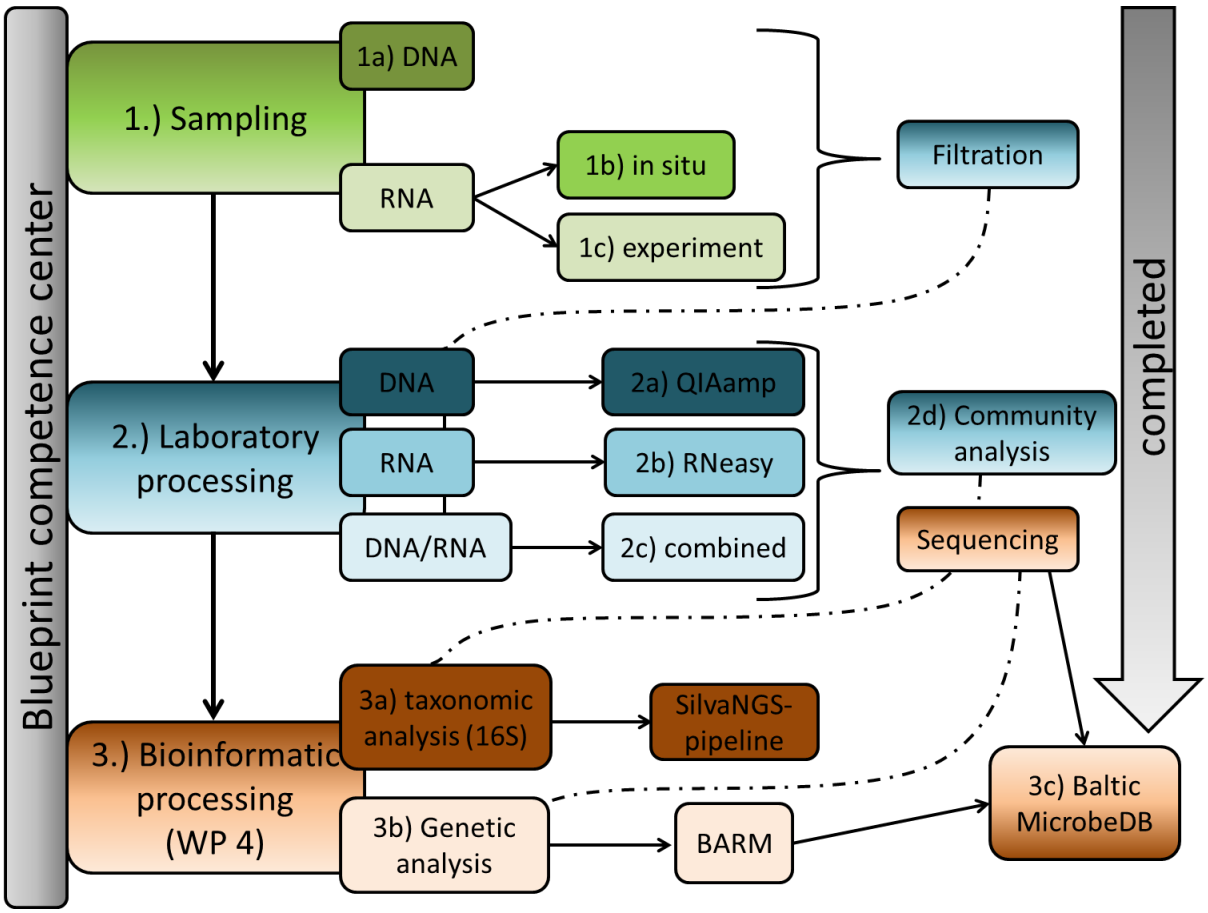

Figure 1: Overview of BCC pipeline from sampling over laboratory processing towards the bioinformatic processing. The establishment of the BalticMicrobe DB is still under development, whereas the sampling and laboratory processing is completed.

In a first step a protocol was established to sample metagenomics and metatranscriptomics independently. Metagenomics are often done to obtain the entire number of genes in an environmental sample and metatranscriptomics give an overview of the activity status of the environment. However, RNA and especially mRNA are very labile and can change within minutes (Steglich et al. 2010) and therefore Feike et al. developed and evaluated an *in situ* sampler (AFIS) where the environmental sample is fixed at the time of sampling (Feike et al.

2012). This method enables unbiased analysis of samples obtained from e. g. anoxic habitats or deeper depths. For their study, Feike et al. developed the AFIS type I sampler, of which the patent was sold to the company HYDROBIOS in Kiel (Germany). HYDROBIOS constructed a new, more secured and fully automatic operating *in situ* sampler (AFIS II). However, the principle of *in situ* fixation stayed the same and is transferable to experimental set ups, where no high-tech equipment is needed. Immediately after sampling the fixative is added and preserves the status quo of the sample.

This is followed by the laboratory procedures regarding community analysis, presence of genes and their expression level. A DNA and RNA extraction protocol was developed using the QIAGEN extraction kits (RNeasy and QIAmp) with modifications. After extraction the material was used for either community analysis using amplicon sequencing by Illumina or send away to companies for metagenome or metatranscriptome sequencing. After sequence acquisition bioinformatic processing was done. At first a Baltic Sea reference metagenome (BARM) was constructed, which is detailed in deliverable 4.1. The BARM is implemented into the BalticMicrobeDB (see deliverable 4.2) where metagenomes and metatranscriptomes from upcoming sampling campaigns are mapped to. The genes within the BARM are functionally and taxonomically annotated and serve as prerequisite for the mapping. At the moment we are determining potential indicator genes. For this the deliverable 3.1 (compilation of signature genes indicative of GES/sub-GES conditions from literature on model microorganisms and natural seawater samples) serves as basic concept.

1.) Sampling

For the sampling it needs to be clarified if size fractionation is needed. Otherwise this step can be omitted.

a) DNA sampling

1. Direct filtration

For the DNA analysis, 1 L of seawater is filtered directly onto a 47 mm Durapore membrane filter (pore size of 0.2 µm; GVWP04700, Merck Millipore, Darmstadt, Germany) using a vacuum of < 300 mbar. Subsequently, the filters are folded, shock frozen using liquid nitrogen and stored at -80°C until further processing.

2. Size fractionation

If size fractionation is required then a pre-filtration step needs to be introduced. Therefore, 1 L of sweater is filtered directly onto a 47 mm Isopore membrane filter (pore size of 3 µm; TSTP04700, Merck Millipore, Darmstadt, Germany) using a vacuum of < 300 mbar. The filtrate is collected in a clean sampling bottle and is further filtered through a 0.2 µm pore-sized Durapore membrane filter (GVWP04700; 47 mm in diameter) using a vacuum of < 300 mbar. Then both filters are folded, shock frozen and stored at -80°C until further processing.

b) RNA sampling *in situ*

1. Direct filtration

Samples for RNA analysis are fixed *in situ* using the AFIS II, which contained a fixative of 5% water-saturated phenol in absolute ethanol (Feike et al. 2012). In the moment of sampling, the AFIS is closed and injects immediately the fixative into the sample in a ratio of 1:10. Similar as for the DNA samples, 1 L of the *in situ* fixed water sample were directly filtered onto 47 mm Durapore membrane filters (GVWP04700, Merck Millipore, Darmstadt, Germany) using a vacuum of < 300 mbar and then processed as described for the DNA analysis. Filtration needs to be done under a fume hood, since the fixative contains phenol.

2. Size fractionation

Also for RNA samples, size fractionation might be needed, therefore 1 L of sweater is filtered directly onto a 47 mm Isopore membrane filter (pore size of 3 µm; TSTP04700, Merck Millipore, Darmstadt, Germany) using a vacuum of < 300 mbar, similar as for DNA samples. The filtrate is collected in a clean sampling bottle and is further filtered through a 0.2 µm pore-sized Durapore membrane filter (GVWP04700; 47 mm in diameter) using a vacuum of < 300 mbar. Then both filters are folded, shock frozen and stored at -80°C until further processing. Filtration needs to be done under a fume hood, since the fixative contains phenol.

c) RNA sampling for experiments

Often, experiments are conducted on a micro- or mesoscale and no *in situ* sampler is needed. However, it is very important for the experiment itself and for the comparability with *in situ* data, that the principle of fixation is adapted to the experimental design. It has been proven to be useful for fixation of experimental samples, to first place the 5% water-saturated phenol in absolute ethanol (Feike et al. 2012) into the sampling bottle and then add the water sample in a ratio of 1:10.

2.) Laboratory processing

a) DNA extraction - QIAamp DNA Mini Kit (Qiagen, cat. No. 51304)

DNA is extracted using a modified protocol of the QIAamp DNA min kit with an initial bead-beating step and a clean-up and concentration process using the Nucleospin gDNA clean-up kit (740230, Machery & Nagel, Dueren, Germany [Figure 2]). The concentration and quality of the eluted DNA were checked by gel electrophoresis and Bioanalyzer DNA 12000 kit (5067-1508, Agilent Technologies, Santa Clara, USA).

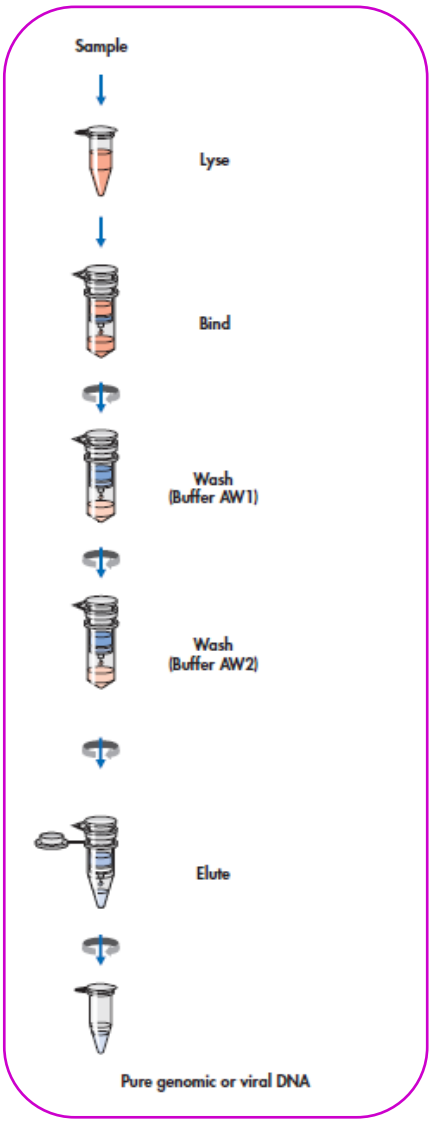

Figure 2: Schematic overview of DNA extraction (source: manual QIAamp DNA mini kit, Qiagen)

1. *Cutting of membrane filter*
   - Cut membrane filters into small pieces and transfer them into a new 2 mL collection tube
2. *Mechanical disruption of cells*
   - Add 100 µL of β-Mercaptoethanol to 10 mL of RLT-Buffer
   - Add 2 scoops of spatula of RNA-free Zirconium beads (200 µm Low Binding Zirconium Beads; BAWZ 200-250-07, OPS Diagnostics LLC, Lebanon, USA) and add 600 µL of RLT-Buffer
   - Optional: add DNA standards (see Satinsky et al. 2013)
   - Vortex (Vortex Genie® 2 Vortex, MO BIO Laboratories, cat. No. 13111-V-220) 5 min at level 7 (~2000 rpm)
   - Centrifuge lysate for 1 min at 10000 g
   - Transfer lysate to new 2 mL collection tube (keep on ice)
   - Add another 600 µL of RLT-Buffer to the remaining beads and vortex 5 min at level 7

- Centrifuge lysate for 1 min at 10000 g
- Transfer lysate to the collection tube kept on ice and centrifuge 2 min at 10000 g
- Transfer lysate into a new 2 mL collection tube (avoid carrying over of pellet)
- Then transfer 700 µL of lysate to DNA spin column and centrifuge for 30 sec at 10000 g
- Discard flow through
- Load remaining lysate onto the same DNA spin column and centrifuge again for 30 sec
- Discard flow through

3. *DNA purification*

- Add 500 µL of AW1 Buffer to DNA spin column
- Centrifuge for 30 sec at 10000 g
- Discard flow through and add 500 µL of AW2 Buffer to spin column
- Centrifuge for 30 sec at 10000 g
- Discard flow through and place spin column into a new 2 mL collection tube and centrifuge for 2 min at 10000 g
- Place spin column into a 1.5 mL collection tube
- Add 100 µL Elution Buffer to column and incubate for 1 min at room temperature
- Then centrifuge for 1 min at 10000 g
- Repeat this step and proceed with quality check

4. *Quality check with PCR*

PCR preparation 15 µL:		Incubation protocol:		
DEPC water	9.62 µL	95°C	4 min	
10 x PCR buffer	1.5 µL	**95°C**	**30 sec**	} x 25 cycles
BSA	0.3 µL	**53°C**	**45 sec**	
$MgCl_2$ (25 mM)	1.5 µL	**72°C**	**45 sec**	
dNTP Mix (10 mM)	0.3 µL	72°C	5 min	
Primer 341f (10 µM)	0.6 µL	12°C	∞	
Primer 805r (10 µM)	0.6 µL			
Taq (DNA 5', 5 u/µL)	0.08 µL			
Template	0.5 µL			

- As template use a 1:100 dilution of the DNA
- Primer Bakt_341F [CCTACGGGNGGCWGCAG]; Primer Bakt_805R [GACTACHVGGGTATCTAATCC] (Herlemann et al. 2011)

5. *Clean-up and concentration of DNA (NucleoSpin® gDNA clean-up, Machery Nagel, cat. No. 740230)*

   - In general the DNA sample is split into two aliquots (2 x 100 µL)
   - Of which one aliquot is stored at -80°C in a special precipitation mix [1.5 mL EtOH molecular grade, 150 µL 2 M NaAc, 15 µL $MgCl_2$] → this is for long-term storage
   - the other aliquot is cleaned and concentrated according to the Nucleospin gDNA instructions (Figure 3)

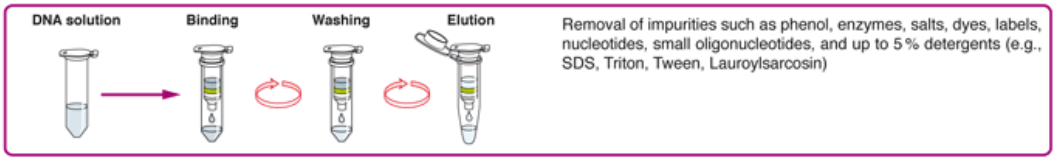

Figure 3: Schematic overview of the clean and concentration process after extraction (source: manual of NucleoSpin® gDNA clean-up from Machery Nagel).

   - adjust the sample volume to 150 µL (therefore add 50 µl RNA/DNA-free water to the eluted DNA) and add 450 µL DB-Buffer and vortex 5 sec
   - load sample on Nucleospin gDNA clean-up column and centrifuge for 30 sec at 11000 g
   - then add 700 µL DW-Buffer to the spin column and vortex 2 sec and then centrifuge at 11000 g for 30 sec, discard the flow-through
   - repeat this step
   - to dry the spin column membrane centrifuge again at 11000 g for 1 min
   - eluting the DNA add 16 µL of DE-Buffer (preheat the DE-Buffer to 70°C) and incubate 1 min at RT
   - then centrifuge for 30 sec at 11000 g
   - repeat this step
   - collect the eluted DNA in a clean reaction tube

6. *Quantification*

   - Quantification and quality check is done using the Bioanalyzer DNA 12000 kit (5067-1508, Agilent Technologies, Santa Clara, USA)

b) RNA extraction - RNeasy Mini kit (Qiagen, cat. No. 74104)

RNA is extracted using a modified protocol of the RNeasy mini kit with an initial bead-beating step and an on-column DNA digestion (79254, Qiagen, Hilden, Germany). The extracted RNA is further subjected to a turbo DNA digestion (AM1907, Fisher Scientific, Schwerte, Germany) after extraction and a final clean-up and concentration process using the Zymo RNA clean and concentrator kit (R1015; Zymo Research Europe, Freiburg, Germany). The concentration and quality of the RNA is checked using the Bioanalyzer RNA 6000 nano kit (5067-1511, Agilent Technologies, Santa Clara, USA).

1. *Cutting of membrane filter*

   - Cut membrane filters into small pieces and transfer them into a new 2 mL collection tube

2. *Mechanical disruption of cells*

   - Add 100 µL of β-Mercaptoethanol to 10 mL of RLT-Buffer
   - Add 2 scoops of spatula of RNA-free Zirconium beads (200 µm Low Binding Zirconium Beads; BAWZ 200-250-07, OPS Diagnostics LLC, Lebanon, USA) and add 600 µL of RLT-Buffer
   - Optional: add mRNA standards (see Satinsky et al. 2013)
   - Vortex (Vortex Genie® 2 Vortex, MO BIO Laboratories, cat. No. 13111-V-220) 5 min at level 7 (~2000 rpm)
   - Centrifuge lysate for 1 min at 10000 g
   - Transfer lysate to new 2 mL collection tube (keep on ice)
   - Add another 600 µL of RLT-Buffer to the remaining beads and vortex 5 min at level 7
   - Centrifuge lysate for 1 min at 10000 g
   - Transfer lysate to the collection tube kept on ice and centrifuge 2 min at 10000 g
   - Transfer lysate into a new 2 mL collection tube (avoid carrying over of pellet)
   - Add 1 volume of 70% EtOH to lysate and mix well by inverting the tube
   - Then transfer 700 µL of lysate to RNA spin column and centrifuge for 30 sec at 10000 g
   - Discard flow through
   - Repeat the loading of RNA spin column until lysate-EtOH mix is used up

3. *On-column DNA digestion (Qiagen, cat. No.* 79254)

   - After binding the RNA to the membrane of the spin column add 500 µL RW1-Buffer to the spin column and centrifuge for 30 sec at 10000 g
   - Discard flow through and place spin column into a new 2 mL collection tube (is not provided by the kit)
   - Prepare DNase mixture: add 10 µL of DNase I stock solution to 70 µL of RDD-Buffer

- Then add these 80 µL of DNase mixture to the spin column and incubate them for 15 min at 30°C (do not shake DNase mixture)

4. *RNA purification*

- Add 400 µL RW1-Buffer to spin column and centrifuge for 30 sec at 10000 g
- Discard flow through and add 500 µL of RPE-Buffer to spin column and centrifuge for 30 sec at 10000 g
- Repeat this step
- Discard flow through and place spin column into a new 2 mL collection tube
- For drying the membrane centrifuge for 2 min at 10000 g
- Place spin column into a new 1.5 mL collection tube
- Add 50 µL of RNase-free water directly to the membrane and incubate for 1 min
- Then centrifuge for 1 min at 10000 g
- Repeat this step and proceed with turbo DNA digestion
- Keep RNA the whole time at 4°C

5. *Turbo DNA digestion (Ambion, cat. No. AM1907)*

- add 10 µl of Dnase reaction buffer to RNA
- then add 3 µl of turbo Dnase and incubate for 20 min at 37°C (do not shake Dnase mix)
- add another 3 µl of turbo Dnase and incubate another 20 min at 37°C
- add 20 µl of the stop solution and vortex up to 1 min on and off
- Afterwards centrifuge 1 min at 13000 g and transfer supernatant to new collection tube avoid carrying-over of milky stop solution

6. *Quality check with PCR*

PCR preparation 15 µL:		Incubation protocol:		
DEPC water	9.62 µL	95°C	4 min	
10 x PCR buffer	1.5 µL	**95°C**	**30 sec**	} x 25 cycles
BSA	0.3 µL	**53°C**	**45 sec**	
$MgCl_2$ (25 mM)	1.5 µL	**72°C**	**45 sec**	
dNTP Mix (10 mM)	0.3 µL	72°C	5 min	
Primer 341f (10 µM)	0.6 µL	12°C	∞	
Primer 805r (10 µM)	0.6 µL			
Taq (DNA 5', 5 u/µL)	0.08 µL			
Template	0.5 µL			

- As template use undiluted RNA extract
- Primer Bakt_341F [CCTACGGGNGGCWGCAG]; Primer Bakt_805R [GACTACHVGGGTATCTAATCC] (Herlemann et al. 2011)

7. *Clean-up and concentration of RNA (cat. No. R1015)*
    - Add 2 volumes of Zymo RNA Binding Buffer to RNA sample
    - Then add equal volume of EtOH (95-100% EtOH molecular grade) and mix well by inverting the tube
    - Transfer mix to Zymo spin column and centrifuge for 30 sec at 13000 g
    - Discard flow through and add 400 µL of RNA Prep Buffer to membrane and centrifuge for 30 sec at 13000 g
    - Discard flow through and add 700 µL of RNA Wash Buffer to membrane and centrifuge for 30 sec at 13000 g
    - Discard flow through and add 400 µL of RNA Wash Buffer to membrane and centrifuge for 30 sec at 13000 g
    - Discard flow through and place spin column in a new 2 mL collection tube and centrifuge for 2 min at 13000 g
    - Transfer spin column to a new 1.5 mL collection tube and add 16 µL RNase-free water directly to the membrane
    - Incubate for 1 min at RT and centrifuge for 1 min at 13000 g
    - Repeat this step and proceed with quantification
    - Keep RNA the whole time at 4°C
8. *Quantification*
    - Quantification and quality check is done using the Bioanalyzer RNA 6000 nano kit (5067-1511)
    - After the Bioanalyzer add 0.9 µL of RiboLock RNase Inhibitor (40 U/L) to the cleaned RNA and store at -80°C

c) <u>simultaneous extraction of DNA and RNA</u>

Often it is required that DNA and RNA needs to be extracted simultaneously from the same membrane filter. Therefore it is possible to combine the extraction protocols for DNA and RNA extractions (Figure 4). The procedure is similar as for the single extraction; however some modifications for the mechanical disruption of cells have to be made, which are listed in the following:

- Add 100 µL of β-Mercaptoethanol to 10 mL of RLT-Buffer
- Add 2 scoops of spatula of RNA-free Zirconium beads (200 µm Low Binding Zirconium Beads; BAWZ 200-250-07, OPS Diagnostics LLC, Lebanon, USA) and add 600 µL of RLT-Buffer
- <u>Optional:</u> add mRNA and DNA standards (see Satinsky et al. 2013)
- Vortex (Vortex Genie® 2 Vortex, MO BIO Laboratories, cat. No. 13111-V-220) 5 min at level 7 (~2000 rpm)
- Centrifuge lysate for 1 min at 10000 g
- Transfer lysate to new 2 mL collection tube (keep on ice)

- Add another 600 µL of RLT-Buffer to the remaining beads and vortex 5 min at level 7
- Centrifuge lysate for 1 min at 10000 g
- Transfer lysate to the collection tube kept on ice and centrifuge 2 min at 10000 g
- Transfer lysate into a new 2 mL collection tube (avoid carrying over of pellet)
- Then transfer 700 µL of lysate to DNA spin column and centrifuge for 30 sec at 10000 g
- Keep flow through and transfer it to a new 2 mL collection tube (RNA, keep on ice)
- Load remaining lysate onto the same DNA spin column and centrifuge again for 30 sec
- Transfer the flow through to the collection tube on ice
- Add 1 volume of 70% EtOH to the RNA lysate and mix well by inverting the tube
- Then transfer 700 µL of lysate to RNA spin column and proceed with the RNA purification (meanwhile leave DNA on ice and proceed with purification after RNA is done)

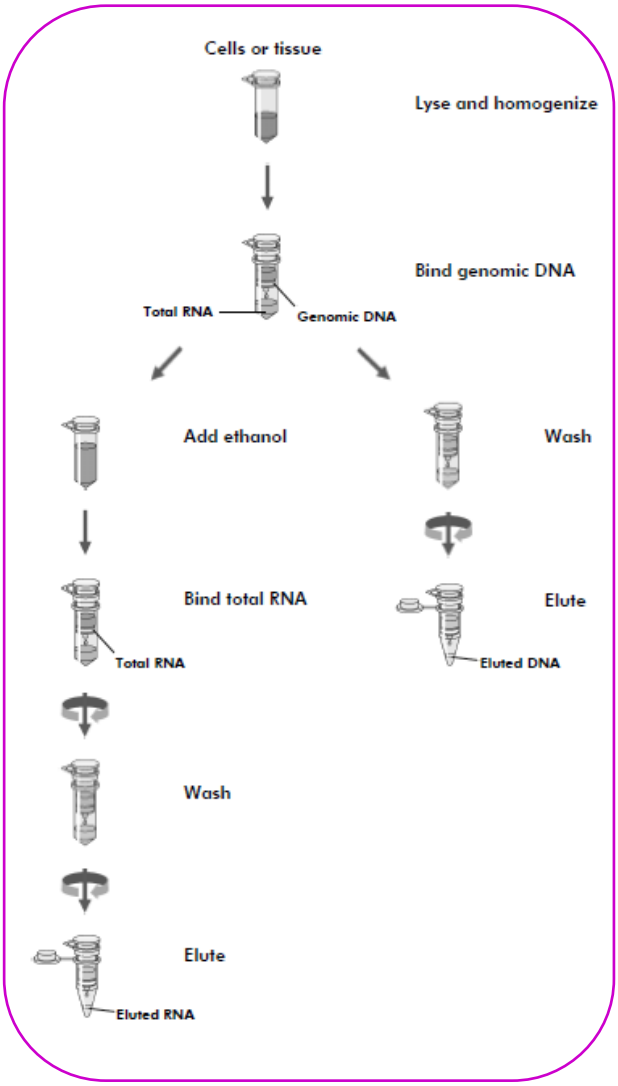

Figure 4: Schematic overview of simultaneous extraction of DNA and RNA from the same sample (source: manual AllPrep mini kit, Qiagen)

d) Community analysis: Amplicon library preparation

The extracted DNA and RNA can be used for community analysis by amplicon sequencing, but the total RNA has to be transcribed into cDNA before preparing the libraries for sequencing. The extracted RNA is transcribed into cDNA using a modified protocol of the TaqMan® reverse transcription kit (4304134, Fisher Scientific, Schwerte, Germany).

1. Reverse Transcription

RT preparation 55 µL:		Incubation protocol:	
DEPC water	Add to 55 µL	25°C	5 min
10 x RT Buffer [50 mM KCl, 10 mM Tris-HCl pH 8.3]	5.5 µL	42°C	45 min
$MgCl_2$ (25 mM)	3.85 µL	85°C	5 min
dNTP Mix (10 mM each)	2.75 µL	10°C	∞
RNase Inhibitor (40 U/L)	1.375 µL		
Multi Scribe RT (50 U/µL)	2.75 µL		
Primer 1492r (10 µM)	2.75 µL		
Template (20 ng)	X µL		

- Primer S-D-Bact-1492-a-A-16 [GGTTACCTTGTTACGACTT] (Alm et al. 1996)

The constructed cDNA is verified by PCR using the primers for the V3/V4 region; for protocol description see quality check.

2. Clean-up of cDNA

The cDNA product is confirmed by gel electrophoresis and then the samples are cleaned using the Agencourt® AMPure® XP kit (A63881, Beckman Coulter, Krefeld, Germany).

- Transfer cDNA products to a 1.5 mL tube
- Add 72 µL of Ampure Reagent for each 50 µL of cDNA product
- Mix well by pipetting and centrifuge quickly (use table centrifuge)
- Place tube into the magnetic bead separator and leave it for 2 min (until solution is clear → beads are on the side)
- Discard supernatant and add 200 µL 70% EtOH
- Repeat this step
- Discard all liquid from tube and leave tube 5 min at 37°C with open lid
- Add 12 µL of 10 mM Tris pH 8.5 and re-suspend everything
- Mix well by pipetting and centrifuge quickly (use table centrifuge)
- Place tube into the magnetic bead separator and leave it for 2 min (until solution is clear → beads are on the side)
- Pipette 10 µl of supernatant and transfer to 0.5 mL tube (or to PCR tube).

The concentrations of the purified cDNA is determined using the Quant-iT™ PicoGreen® dsDNA assay (P11496, Fisher Scientific, Schwerte, Germany) before they are used as templates for the amplicon PCR.

3. Amplicon library preparation

Amplicon PCR preparation 25 µL:

2 x KAPA Hifi Hot Start Ready Mix (KK2601)	12.5 µL
Amplicon Primer (1 µM)	5 µL
Amplicon Primer (1 µM)	5 µL
cDNA (5 ng/µL)	2.5 µL

Incubation protocol:

95°C 3 min
95°C 30 sec
55°C 30 sec } x 25 cycles
72°C 30 sec
72°C 5 min
4°C ∞

- The V3/V4 region of the 16S rRNA carry the Illumina adapters
- Bakt_341F:TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG
- Bakt_805R:GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTA TCTAATCC
- The amplicon PCR products are cleaned according to the instructions included with the Agencourt® AMPure® XP kit and quality checked using the Bioanalyzer 1000 DNA kit.
- The expected size of the PCR product obtained with the V3/V4 primer pair was ~550 bp.

4. Indexing of amplicon libraries

Dual indices, which function as an identification marker, need to be attached to the amplicon PCR product as follows:

Index PCR preparation 25 µL**:**

2 x KAPA Hifi Hot Start Ready Mix (KK2601)	12.5 µL
Nextera XT Index 1 Primers	2.5 µL
Nextera XT Index 2 Primers	5 µL
cDNA (cleaned Amplicon PCR product)	2.5 µL

Incubation protocol:

95°C 3 min
95°C 30 sec
55°C 30 sec } x 8 cycles
72°C 30 sec
72°C 5 min
4°C ∞

- The amplicon PCR products are cleaned according to the instructions included with the Agencourt® AMPure® XP kit

- For quality control and determination of the DNA concentration, a 1:10 dilution of the template DNA was prepared and checked using the Bioanalyzer 1000 DNA kit
- The expected size of the PCR product is ~630 bp.

5. Pooling and sequencing

The libraries are then pooled after their dilution to obtain a concentration of 4 nM. From each pooled library, 4 pM are Illumina sequenced using the Illumina 600 v3 reagent cycle kit and a paired end protocol. The PhiX control is spiked into each of the pools to a final amount of 10%. A cluster density of ~1200 (K mm^{-2}) needs to be achieved in over 80% of the sequencing and index reads with a Q-score ≥ 30.

3.) Bioinformatic processing

a) Taxonomic analysis

After sequencing the amplicon libraries as paired ends, the single forward and reverse Illumina reads are joined using the fastq-join script (join_paired_ends.py) by (Aronesty 2013) with a minimum overlap of 20 bp. This will result in three fastq-formatted output files: '*.join' (assembled/joined reads), '*.un1' (unassembled forward reads) and '*.un2' (unassembled reverse reads). For the downstream analysis, only the '*.join' reads are considered for processing; the '*.un1' and '*.un2' reads are discarded.

For further analysis the SILVAngs-pipeline (https://www.arb-silva.de/ngs/) is a very suitable web-based platform for taxonomic analysis. However, to use this pipeline the fastq-formatted files need to be transformed into fasta-formatted files as a perquisite for the analysis (Quast et al. 2013). The SILVAngs-pipeline aligns all reads using SINA (Pruesse et al. 2012); further quality filtering, including determinations of length, ambiguity, and the presence of homopolymers, is done automatically. Only reads with sufficient quality values are used in the de-replication step. The reads are de-replicated at a 100% identity level and subsequently clustered at 98% using cd-hit-est (version 3.1.2; http://www.bioinformatics.org/cd-hit; (Li & Godzik 2006). Representative OTUs (operational taxonomic unit) are classified to the genus level in accordance with the non-redundant version of the SILVA SSU Ref dataset (Quast et al. 2013) using blastn (version 2.2.28+; http://blast.ncbi.nlm.nih.gov/Blast.cgi) with a similarity threshold of 93% (Camacho et al. 2009). The classification of each OTU reference read is mapped onto all reads assigned to the respective OTU. This will yield quantitative information (number of individual reads per taxonomic path) within the limitations of the PCR and sequencing technique biases and takes into account multiple rRNA operons. Reads, which do not exceed the similarity threshold of 93%, are classified as 'No Relative'.

b) Genetic analysis

The bioinformatic processing for metagenomic and metatranscriptomic data is organized in WP4 of the BLUEPRINT project and has been described in detail in deliverables 4.1 and 4.2. Here, we will give a brief summary.

First a Baltic Sea reference metagenome was constructed and new samples (metagenomic and/or metatranscriptomics) can be mapped on to identify blueprints (Figure 5).

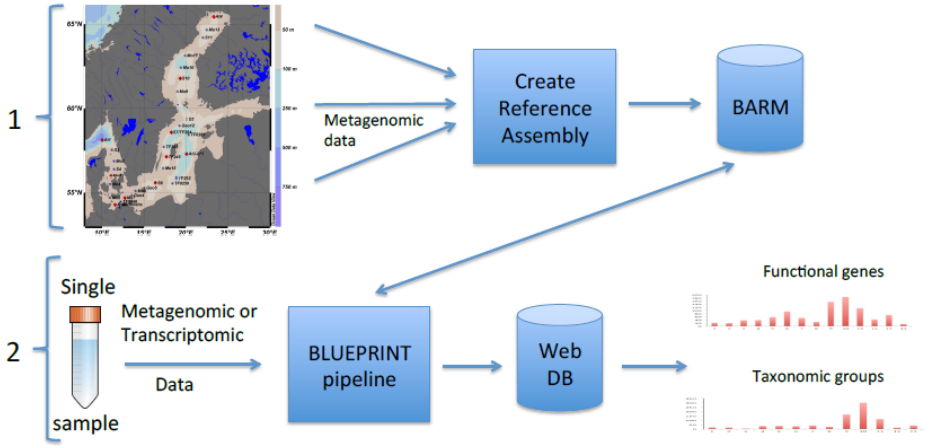

Figure 5: Schematic overview of the bioinformatic workflow in the BONUS BLUEPRINT project. (1) Assembly and annotation of the Baltic Sea reference metagenome (BARM). (2) Obtaining blueprints from samples. (source: Deliverable 4.1, WP4, BLUEPRINT project).

For the mapping it is very important that the obtained sequences of the specific sample are quality trimmed before they are mapped onto the BARM. Mapping is done using the program Bowtie2 (Langmead et al. 2012) and removal of redundant reads is done by the program Picard (http://broadinstitute.github.io/picard/). Those reads which finally mapped to each gene on the contigs are counted and normalized by the length of the gene and by the total number of reads in the sample using RPKM value (reads per kilobase of gene sequence of gene sequence per million reads mapped). A RPKM value is calculated for every gene in the BARM. A bioinformatic pipeline and the BalticMicrobeDB have been set up (http://github.com/EnvGen/BLUEPRINT_pipeline) and each functional and taxonomic annotation is associated to each gene in the BARM.

Costs

One of the responsibilities of the BCC is to establish low cost sampling and laboratory protocols. Figure 6 shows the costs per sample in Euro for each step in the BCC pipeline. Users have to decide what kind of sample they want to process and what kind of information is suitable to answer their specific question. For example, if taxonomic composition is the appropriate indicator the approximate costs per sample are 80 €. However, if information on expressed genes (level of activity) is required the total costs per sample are 350 €. Prices do not contain personal and equipment costs, since these can vary between laboratories. Prices for sequencing are without tax and were obtained from LGC Genomics, which gives currently the best quotes for sequencing. However, in future prices for sequencing will likely decrease further, which makes the Blueprint idea using indicator genes for state description even more approachable.

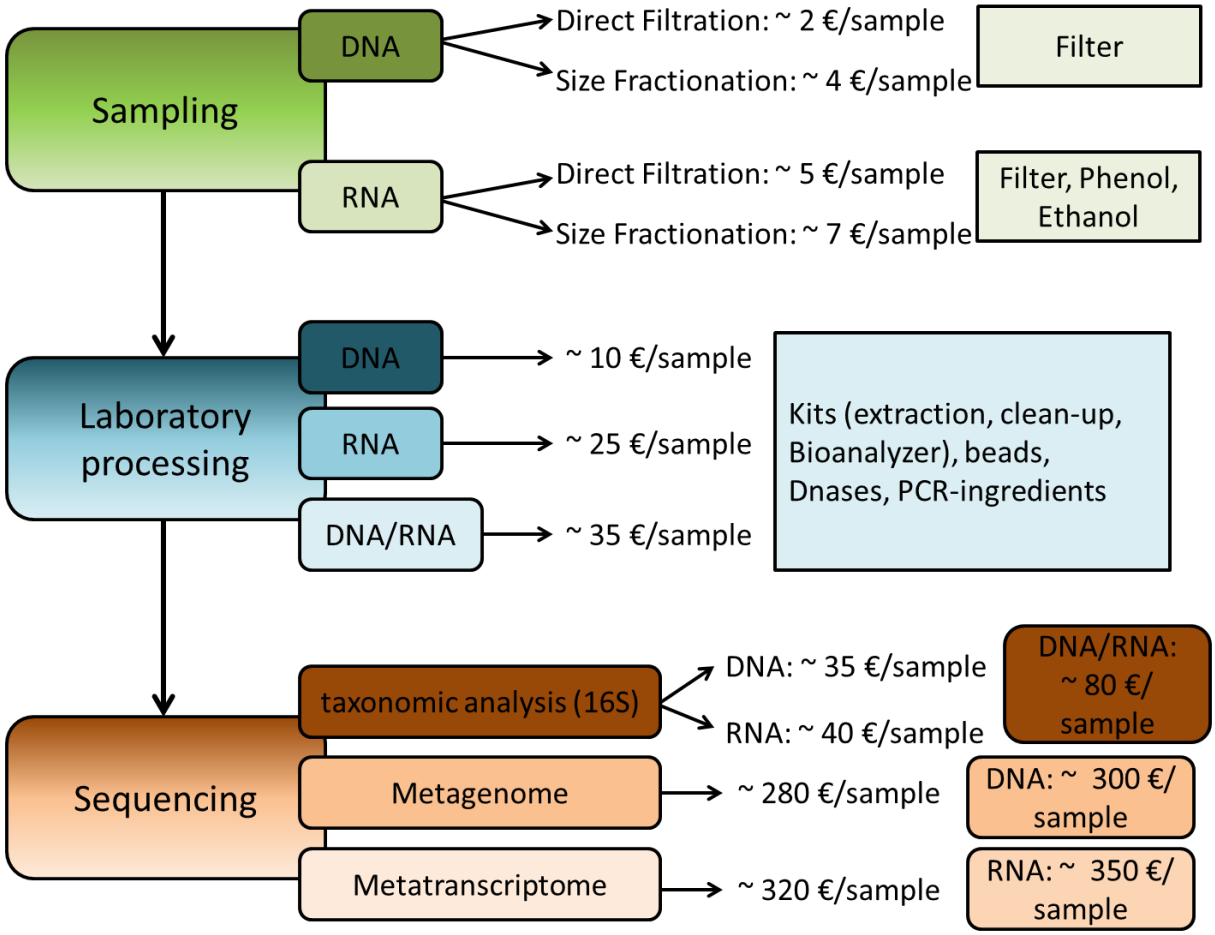

Figure 6: Overview of costs per sample in EURO. Prices include all consumables but no equipment and personal costs and are given separately for each step within in the Blueprint pipeline. Sequencing is usually carried out by companies and prices therefore where obtained from LGC Genomics.

Conclusion

In summary, the sampling protocol is established and the laboratory processing of the samples is completed. The established protocols have successfully been used for processing samples from two Baltic Sea cruises in June 2014 (AL439) and August 2015 (POS488) and in several laboratories of the Blueprint partners (WP 1, 2, 3 and 6). The protocol was also adopted by the BONUS AFISmon project. The manuscript "The use of a general 16S rRNA primer system to identify phytoplankton throughout the Baltic Sea" by Bennke, C. M., Pollehne, F., Müller, A., Hansen, R., Kreikemeyer, B. and Labrenz, M. is submitted for publication at L&O Methods, where the sampling and laboratory processing as well the community analysis is described in detail. Currently, a manuscript regarding the genetic analysis of metatranscriptomic data is in progress.

References:

Alm, E. W., Oerther, D. B., Larsen, N., Stahl, D. A. & Raskin, L. 1996. The oligonucleotide probe database. Applied and Environmental Microbiology. 10, p. 3557–3559.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. & Madden, T. L. 2009. BLAST+: architecture and applications. BMC Bioinformatics, p. 421.

Feike, J., Jurgens, K., Hollibaugh, J. T., Kruger, S., Jost, G. & Labrenz, M. 2012. Measuring unbiased metatranscriptomics in suboxic waters of the central Baltic Sea using a new *in situ* fixation system. The ISME Journal. 2, p. 461–470.

Herlemann, D. P., Labrenz, M., Jurgens, K., Bertilsson, S., Waniek, J. J. & Andersson, A. F. 2011. Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. The ISME Journal. 10, p. 1571–1579.

Langmead, B. & Salzberg, S. L. 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods. 9, p. 357-359.

Li, W. & Godzik, A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 13, p. 1658–1659.

Pruesse, E., Peplies, J. & Glockner, F. O. 2012. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics. 14, p. 1823–1829.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. & Glockner, F. O. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Research. Database issue, p. D590-6.

Satinsky, B. M., Gifford, S. M., Crump, B. C. & Moran, M. A. 2013: Use of internal standards for quantitative metatranscriptome and metagenome analysis. Methods in Enzymology. 531, p. 237-250.

Steglich C, Lindell D, Futschik M, Rector T, Steen R, Chisholm S. 2010. Short RNA half-lives in the slow growing marine cyanobacterium *Prochlorococcus*. Genome Biology 11, p. R54.