

TARTU ÜLIKOOL  
LOODUS- JA TÄPPISTEADUSTE VALDKOND  
MOLEKULAAR- JA RAKUBIOLOOGIA INSTITUUT  
ÖKOLOOGIA JA MAATEADUSTE INSTITUUT

Keiju Rootsma  
**Mullaorganismide metatranskriptoomi analüüs**  
Bakalaureusetöö

Juhendaja PhD Leho Tedersoo

TARTU 2019

# INFOLEHT

## Mullaorganismide metatranskriptoomi analüüs

Metatranskriptomika on mikrobioloogiline molekulaarne meetod, mis keskendub bioloogilistes proovides leiduvate mikroobide aktiivsuse analüüsimisele ning on efektiivne viis nii prokarüootide kui ka eukarüootide uurimiseks. Käesolev bakalaureusetöö põhineb mullas leiduvate organismide ja nendes toimuvate protsesside tuvastamisel. Teoreetilises pooles antakse ülevaade metatranskriptomika tõhususest ja puudustest võrrelduna teiste sarnaste meetoditega. Praktiline osa keskendub RNA raamatukogude valmistamisele, sünteesimisele ning saadud tulemuste analüüsimisele.

**Märksõnad:** mikroobid, prokarüoodid, eukarüoodid, RNA raamatukogu, metatranskriptom

**B230** mikrobioloogia, bakterioloogia, viroloogia, mükoloogia

## Metatranscriptomic analysis of soil organisms

Metatranscriptomics is a microbiological research method that is used for the analysis of microbial activity within biological samples based on the transcribed RNA. This bachelor's thesis focuses on soil organisms and their genetically expressed activities. The theoretical part of the work gives an overview about the advantages and disadvantages of metatranscriptomics compared to other similar methods. The practical part handles RNA library preparation, cDNA sequencing and bioinformatics analysis of sequence data.

**Keywords:** microbes, prokaryotes, eukaryotes, RNA library, metatranscriptomics

**B230** microbiology, bacteriology, virology, mycology

# SISUKORD

INFOLEHT.....	2
SISUKORD .....	3
KASUTATUD LÜHENDID .....	4
SISSEJUHATUS .....	6
1. KIRJANDUSE ÜLEVAADE.....	7
1.1 Mulla mikroorganismid.....	7
1.1.1 Bakterite ökoloogiline roll .....	7
1.1.2 Seente ökoloogiline roll .....	8
1.2 Esimese ja teise põlvkonna sekveneerimismeetodid .....	8
1.3 Metagenoomika.....	9
1.4 Metatranskriptoomika .....	10
1.5 Metatranskriptoomi analüüsimeetodid ja tarkvara.....	12
1.5.1 RNA terviklikkus ja RIN väärtus.....	12
1.5.2 Analüüsikeskkonnad ja andmebaasid .....	14
2. EKSPERIMENTAALOSA .....	15
2.1 Töö eesmärgid.....	15
2.2 Materjal ja meetodika .....	15
2.2.1 Proovialad ja töötused.....	15
2.2.2 Fenool-kloroform lahuse valmistamine .....	17
2.2.3 RNA ja DNA eraldus .....	17
2.2.4 DNase töötus .....	18
2.2.5 RNA kvaliteedikontroll.....	18
2.2.6 cDNA süntees ja kontsentratsioonide mõõtmine .....	19
2.2.7 RNA raamatukogud .....	20
2.2.8 Bioinformaatiline analüüs .....	22
2.2.9 Statistiline analüüs .....	23
2.3 Tulemused ja arutelu.....	23
2.3.1 Esimese projekti proovide erinevate säilitusmeetodite analüüs.....	23
2.3.2 RNA raamatukogude analüüs .....	25
KOKKUVÕTE .....	31
KASUTATUD KIRJANDUS.....	33
KASUTATUD VEEBIAADRESSID.....	38
LISA 1 .....	39
Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks.....	41

## KASUTATUD LÜHENDID

MG-RAST	metagenoomika analüüsiserver ( <i>metagenomics analysis server</i> )
metaSPAdes	metagenoomiliste andmete konveier ( <i>a pipeline for metagenomic data sets</i> )
cDNA	komplementaarne DNA ahel ( <i>complementary DNA</i> )
ssDNA	üheaheelaline DNA ahel ( <i>single-stranded DNA</i> )
dsDNA	kaheaheelaline DNA ahel ( <i>double-stranded DNA</i> )
RIN	RNA intaktsusnumber ( <i>RNA integrity number</i> )
LN	vedel lämmastik ( <i>liquid nitrogen</i> )
TCO <sub>2</sub>	kogu-süsihappegaas ( <i>total carbon dioxide</i> )
OTU	taksonoomiline üksus ( <i>operational taxonomic unit</i> )
mRNA	informatsiooni-RNA ( <i>messenger-RNA</i> )
PCR	polümeraasi ahelreaktsioon ( <i>polymerase chain reaction</i> )
NGS	uue-generatsiooni sekveneerimine ( <i>next-generation sequencing</i> )
RT-PCR	pöördtranskriptsiooni polümeraasi ahelreaktsioon ( <i>reverse transcription polymerase chain reaction</i> )
RNA QC	RNA sekveneerimistulemuste kvaliteedikontrolli meetod ( <i>RNA-Seq data quality control method</i> )
rRNA	ribosomaalne RNA ( <i>ribosomal RNA</i> )
<i>de novo</i>	ilma referentsgenoomita transkriptoom ( <i>creating a transcriptome without the help of a reference genome</i> )

COG	valkude funktsiooni andmebaas ( <i>Clusters of Orthologous Groups of proteins</i> )
eggNOG	ortoloogsete rühmade andmebaas ( <i>Non-supervised Orthologous Groups</i> )
KEGG	geenide ja genoomide nomenklatuuri entsüklopeedia ( <i>Kyoto Encyclopedia of Genes and Genomes</i> )
SEED	<i>de novo</i> annotatsioonide platvorm ( <i>a platform for discovering de novo annotations</i> )
IGM/M	metagenoomika andmebaas ( <i>The Integrated Microbial Genomes with Microbiome Samples</i> )
ITS	sisemine transkribeeritav speisser ( <i>internal transcribed spacer</i> )
bp	aluspaar ( <i>base pair</i> )

## SISSEJUHATUS

Muld on kompleksne keskkond, milles elab miljoneid keerulistes toiduvõrgustikes osalevaid baktereid, arhesid ning eukarüoote. Prokarüoodid (bakterid ja arhed) on pika evolutsioneerumise käigus saavutanud erakordse mitmekesisuse ning moodustavad väga olulise osa biosfäärist. Suurem osa neist on veel tänapäevalgi avastamata. (Torsvik jt., 2002). Mullas elavad eukarüootsed mikroorganismid on teatud ökosüsteemides oluline osa kogu mikroobide biomassist. Enamikku mikroobe, eriti arhesid, on keeruline pinnasest isoleerida ja laboris katseklaasi meetodil kasvatada. (Bailly jt., 2007).

Metatranskriptomika ühendab omavahel keemia ning bioloogia ja võimaldab avastada suurt osa seni tundmatust mulla mikrobiomist. (Handelsman jt., 1998). Metatranskriptomika vaatleb konkreetsel ajahetkel transkribeeritud geene ja nende arvukust. Integreerides metatranskriptomika teiste paralleelsete uuringutega, näiteks sihtmärk-geeni sekveneerimise või metagenoomikaga, on võimalik läbi viia väga suure potentsiaaliga teadustööd, mis võimaldavad vaadelda erinevate organismide vahelisi interaktsioone, kommunikatsiooni ja nende metabolismi. Erinevatel ajahetkedel tehtud metatranskriptomiliste uuringute võrdlemine annab teavet kogu mikroorganismide metaboolse dünaamika kohta, mis on vastavuses pidevalt muutuvate keskkonnatingimustega. Transkriptomilised ja metatranskriptomilised uuringud hõlbustavad ka uute ensüümide ja ökofüsioloogiliste protsesside avastamist, mis on olulised nii bioenergeetikas kui ka meditsiinis, põllumajanduses ja keemiatööstuses. (Kuske jt., 2014).

Käesoleva bakalaureusetöö raames tegeleti korraga kahe metatranskriptomika projektiga. Esimene projekt oli erinevate säilitusmeetoditega hoiustatud bioloogiliste proovide metatranskriptomiline analüüs. Selle eesmärgiks oli tuvastada võimalikud viisid kvaliteetse RNA säilitamiseks ja vaadelda, kuidas erinevatest säilitusmeetoditest tulenev proovide kvaliteet RIN väärtusena peegeldab geeniekspressiooni analüüsi tulemusi. Lisaks ka identifitseerida proovides leiduvad organismid ning nende funktsioonid vaadeldes analüüsikeskkondade MG-RAST ja BigTableViewerLite tõhusust. Teine projekt oli Dr. Ingrid Liivi pooleli jäänud Agali ülejutuskatsetes kogutud mulla metatranskriptoomi analüüs, mille eesmärgiks oli identifitseerida proovides olevad organismid ja nende funktsioonid mullas.

Bakalaureusetöö viidi läbi Tartu Ülikooli Ökoloogia ja maateaduste instituudi mükoloogia laboris.

# 1. KIRJANDUSE ÜLEVAADE

## 1.1 Mulla mikroorganismid

Muld on ökoloogiliselt oluline bioloogilist mitmekesisust edendav geenide ja mikroobikoosluste kogum. (<https://www.eea.europa.eu/et/themes/soil/intro>). Selle põhilised komponendid on mikroorganismid, mineraalained ja orgaanilised ühendid, mis on omavahel tihedalt seotud ja mõjutavad maismaa-ökosüsteemide keemilisi, füüsikalisi ja bioloogilisi omadusi ning protsesse. (Mohammadi jt., 2011). Mullas elavad organismid on funktsionaalselt äärmiselt olulised, osaledes lämmastiku- ja süsinikuringes, ökosüsteemi reageeringus globaalsetele muutustele ja taimekoosluste dünaamilisuses. (Bardgett jt., 2014). Mulla mikroorganismide aktiivsus aitab mõista keskkonna ja geneetika mõju taimede kasvule ja ökosüsteemi funktsioneerimisele. (Manter jt., 2017). Enne molekulaarsete meetodite arengut oli teadlastel võimalik tuvastada ainult mõni protsent mikroobidest – bakteritest, arhedest, seentest ja protistidest, mida oli võimalik laboratooriumis kultiveerida. Tänu määramismeetodite arenemisele on nüüdseks teada, et mullas elab kuni 10 miljonit bakteriliiki, 6 miljonit seeneliiki, sadu tuhandeid ümarussiliike ja ka ainurakseid üle maailma. (Carey, 2016).

### 1.1.1 Bakterite ökoloogiline roll

Arvatakse, et muld on kõige suurema elurikkusega bakterikoosluste elupaik. (Roesch jt., 2007). Bakterid on elu edendavate keskkonna-alaste protsesside vahendajad ja seega on bakterite bioloogilise mitmekesisuse mõistmine oluline. (Horner-Devine jt., 2004). Bakterite ökoloogilise rolli mõistmine aitab aru saada, kuidas mullabakterite kooslused ajas ja paiguti erinevad ja inimtekkelistele muutustele reageerivad. (Delgado-Baquerizo jt., 2018).

Mullas elavate bakterite klassifitseerimiseks töötasid Noah Fierer ja tema kolleegid välja oligotroofide-mesotroofide teooria, milles analüüsiti mulla pH taset, niiskust, süsinikusisaldust, tekstuuri, lämmastikusisaldust ning lämmastiku ja süsiniku omavahelist suhet. Süsiniku kättesaadavus mikroorganismidele osutus parimaks mikroobse biomassi indikaatoriks mullas. Mesotroofsed bakterid, näiteks liigid rühmadest *β-Proteobacteria* ja *Bacteroidetes*, elavad energia- ja toitainerikas keskkonnas, samas kui oligotroofsed bakterid, näiteks *Acidobacteria* liigid, elavad pigem vaeses keskkonnas. (Fierer jt., 2007).

### 1.1.2 Seente ökoloogiline roll

Mullas elavad seeneliigid interakteeruvad mitmesuguste bakterite ja eukarüootidega, reguleerides süsiniku ja toitainete ringlust. Seened on funktsionaalselt ühed kõige mitmekesisemad organismid, omades ökoloogiliselt olulist rolli kui orgaanilise aine põhilised lagundajad ekstratsellulaarsete ensüümide eritamise kaudu. Üheskoos taimedega moodustavad seened läbi mükoriisa ning patoloogiliste ja endofüütsete ühenduste metaboolseid sidemeid. Mükoriisaseened vastutavad ka taimede mineraalide kättesaadavuse eest. (Kuske jt., 2014). Seente elu mullas mõjutavad mulla pH tase, temperatuur, hapniku kättesaadavus ja niiskus. Happelistes muldades on seente biomass kõrgem, aga elurikkus madal. Neutraalse pH tasemega muldades on seevastu seente biomass madalam, ent liigirikkus kõrgem. (Kutateladze jt., 2016).

Seente liigirikkus ja funktsioonid mullas sõltuvad ka aastaajast. Suvel korjatud bioloogilistes proovides on seene transkriptide arvukus suurem ning domineerivad liigid *Actinobacteria*, *Proteobacteria* ja *Planctomycetes*. Talve proovides on seente arvukus seevastu väiksem ja põhilised liigid, mida leida võib on *Bacteroidetes* ja *Chlorobi*. Funktsionaalselt on sõltuvalt aastaajast erinev ka valiini, leutsiini ja isoleutsiini lagundamine ja fenüülalaniini metabolism, mille protsentuaalne kasv toimub just suvel. (Žifčáková jt., 2016).

## 1.2 Esimese ja teise põlvkonna sekveneerimismeetodid

Sangeri tehnoloogiat ehk esimese põlvkonna sekveneerimismeetodit tutvustati esmakordselt 1977. aastal avaldatud artiklis, kus Frederick Sanger kolleegidega kirjeldas uusimat meetodit DNA-s nukleotiidide järjestuse määramiseks DNA polümeraas I abil. (Sanger jt., 1977). Sangeri sekveneerimismeetodit kasutati Inimese genoomi projektis (The Human Genome Project), mis osutus väga kalliks. (Collins jt., 2003). Suure kulukuse ja tohutu ajamahu tõttu töötati välja uue generatsiooni sekveneerimismeetodid (NGS) eesmärgiga saada odavamad, suurema läbilaskevõime ja täpsusega tehnoloogiad. NGS kasutuselevõtmine vähendas märgatavalt sekveneerimise hinda, mis miljoni aluspaari kohta on Sangeri meetodi puhul 2400 dollarit, tänapäeval Hiseq 2000 puhul vaid 0,07 dollarit. (Liu jt., 2012). NGS kasutatakse ka metatranskriptoomikas (Lee jt., 2013) kuigi esimesed massiivsed tööd ilmusid Sangeri meetodit kasutades. Kõrge läbilaskevõimega NGS sekveneerimismeetodid nagu Illumina HiSeq, MiSeq ja NovaSeq sekveneerimine pakuvad kõrget taksonoomilist ja funktsionaalset katvust ning parandavad arusaama mikroorganismide ökoloogiast. (Tedersoo jt., 2015). Need väljastavad lühikesi, 300 bp (aluspaari) pikkusi järjestusi väga kõrgel läbilaskevõimel (20, 300 ja 6000 GB) ja lisaks pakuvad fragmendi mõlema otsa sekveneerimisel kõrgkvaliteetsete sekveneerimistulemuste saamist. (Nilsson jt., 2019). Metagenoomika ja metatranskriptoomika

on meetodid, mille abil on võimalik mõista taimede ja mikroobide vahelisi interaktsioone ning kirjeldada vastavalt kogu koosluse genoome ja geenide ekspressiooni. Need meetodid on kõige efektiivsemad kombineeritult. (Carvalhais jt., 2013).

### 1.3 Metagenoomika

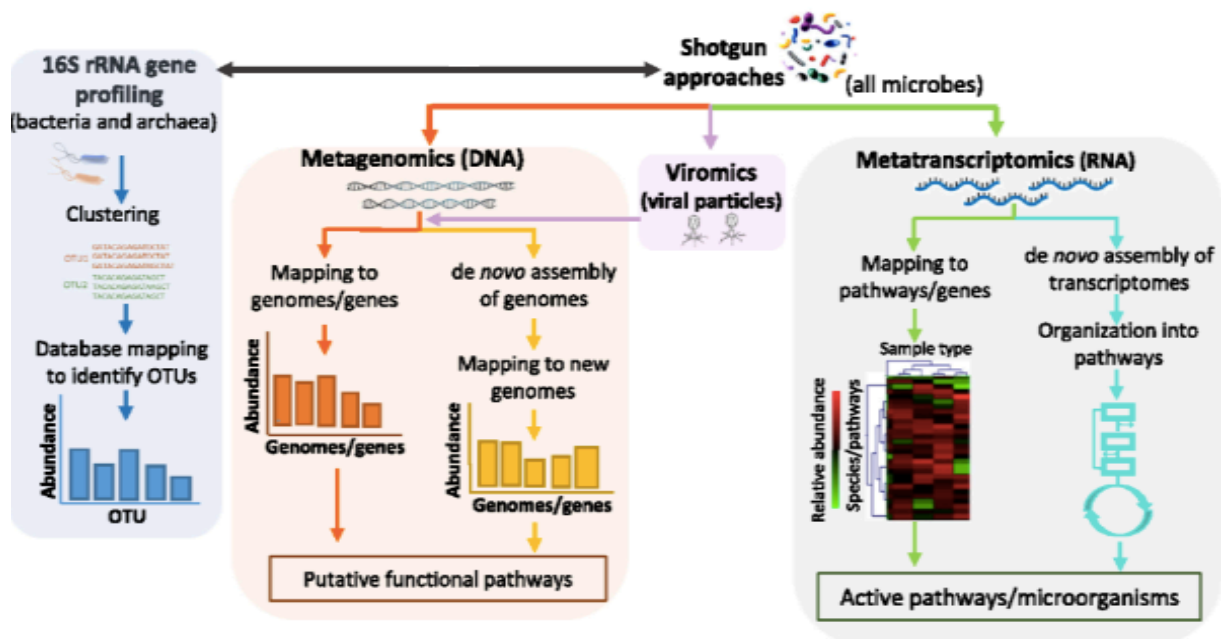
Metagenoomika on meetod, mille abil on võimalik välja selgitada kõigi keskkonnas olevate aktiivsete ja inaktiivsete organismide genoomsed kodeerivad ja mittekodeerivad järjestused, mille põhjal saab teha järeldusi organismide potentsiaalse funktsiooni ja taksonoomilise kuuluvuse kohta. (Aguiar-Pulido jt., 2016). Metagenoomika eelis markergeenide sekveneerimise ees on võime tuvastada organisme nende liigi või tüve tasemeni. Kogu koosluse DNA sekveneerimine annab osalise ülevaate mikroobikoosluse funktsionaalsusest, aga ei paku informatsiooni terves koosluses ekspresseeritud geenide kohta. Seevastu metatranskriptoomika põhirõhk on küsimusel: „Milliseid gene ekspresseeriti kollektiivselt erinevates keskkonnatingimustes?“. Metatranskriptoomi sekveneerimine toob välja organismid, mis on aktiivsed. (Cao jt., 2017; Aguiar-Pulido jt., 2016).

Petr Baldrian ja kolleegid (2012) ühendasid mikroobikoosluste uurimiseks 16S rRNA geenijärjestused ja RNA sekveneerimisel saadud 16S rRNA transkriptide järjestused bakterite määramiseks ning rRNA geenide vahel asuvad sisemiselt transkribeeritavaid speisser (ITS) regioonid nii DNA- kui ka RNA-põhiselt seente määramiseks (seentel ei ole ainuüksi rRNA geenid liikide eristamiseks piisavad, vaid selleks kasutatakse ITS järjestusi; Nilsson jt., 2019). Seente ITS järjestusi on mullas 1000-10000 korda vähem kui bakterite 16S rRNA geenide järjestusi (Baldrian jt., 2012; Tedersoo jt., 2015). Selgus, et mitmed seente ja bakterite OTU-d olid unikaalsed DNA-põhises analüüsis, ent teised esinesid ainult RNA-põhises analüüsis. See näitab, et DNA-põhistel uuringutel jääb teatud osa mikroobikooslusest tuvastamata. (Baldrian jt., 2012). Metagenoomika limiteerivaks faktoriks on ka analüüsimeetodid, mis toetuvad osaliselt varem sekveneeritud genoomidest pärit infole ja tulemused on seetõttu mõjutatud konkreetsete sekveneeritud genoomide poolt. Samuti on massiline metagenoomide sekveneerimine veel üsna kallis. (Quince jt., 2017).

## 1.4 Metatranskriptomika

Arvatakse, et mikroobikoosluste põhjalikumaks uurimiseks on vaja metagenoomika ühendada näiteks metatranskriptomikaga. (Aguiar-Pulido jt., 2016). Metatranskriptomiliste analüüside meetod põhineb kogu RNA-st valmistatud cDNA sekveneerimisel. See on hindamatu meetod koosluses olevate metaboolselt aktiivsete mikroobide eristamisel. (Cao jt., 2017). Metatranskriptomika annab võimaluse tuvastada kõik elusad mikroorganismid, sealhulgas bakterid, arhed ja viirused ning selgitada välja nende funktsioonid ning geeniekspressiooni tasemed kindlatel ajahetkedel ja keskkonnatingimustel. (Toseland jt., 2014). See võimaldab ka mullas elavate protistide mitmekesisust uurida, mida seni on tehtud vähem kui prokarüootide ja eukarüootide puhul. (Geisen jt., 2015). Sellel on suur potentsiaal bioloogilise aktiivsuse informatsiooni avastamiseks, mida teised genoomilised uuringud ei võimalda. Metatranskriptomika näitab ka millised regulatoorsed mehhanismid geeniekspressiooni reguleerivad ehk kuidas peremees-mikroobi ja mikroob-mikroobi omavaheline interakteerumine muudavad mikrobioomi terviklikkust. (Bashiardes jt., 2016).

Interaktsioonide tuvastamiseks metatranskriptomide põhjal tuleb arvestada eksperimentide õige kavandamisega. Üks võimalustest on vaadelda võimalikult palju parameetreid korraga – nii metagenoomi, metatranskriptoomi, keskkonnafaktoreid, metaproteoomi ja metabooloomi – ja seejärel uurida tekkinud transkriptsiooniliste mustrite vastust mõõdetud muutujatele. Sel viisil on raske tuvastada, millised mõõdetud muutujatest tõesti transkriptsiooni mõjutavad. Alternatiivne viis on kontrollitud katsete läbi viimine, mis hõlbustab geenide ja geneetiliste radade tuvastamist. (Carvalhais jt., 2013). Metatranskriptomilistes võrdlusandmebaasides on aga puudus kõrge kvaliteediga, korrektselt annoteeritud ja fülogeneetiliselt mitmekesisest genoomidest, mistõttu on transkriptide ja metatranskriptomiliste andmete tõlgendamine limiteeritud. (Kuske jt., 2014). Samuti on metatranskriptomii limiteerivaks faktoriks proovidest ribosomaalse RNA (rRNA) eraldamine, mis ei pruugi alati õnnestuda. rRNA suur hulk segab transkriptomilisi uuringuid vähendades informatsiooni RNA (mRNA) hulka proovides. Teine probleem on mRNA ebastabiilsus, mis võib tekitada sekveneerimisele eelnevalt muudatusi RNA kvaliteedis. (<https://www.cd-genomics.com/the-principles-workflow-and-applications-of-metatranscriptomic-sequencing.html>).



**Joonis 1: Metatriipkoodistamise (vasakul), metagenoomika (keskel) ja metatranskriptoomika (paremal) võrdlus.** Saadud toorjärjestused läbivad kõigepealt kvaliteedifiltrid. Bakterite puhul kasutatakse taksonite määramiseks peamiselt 16S rRNA geenide järjestusi, mida saadakse kas PCR kasutades või metagenoomist 16S geeni järjestuste eemaldamises, mille järel sekventsid klasterdatakse OTUdeks, mis esindavad sarnaseid organisme ja määratakse igale OTU-le taksonoomiline kuuluvus, mis põhineb saadud järjestuse ja 16S rRNA andmebaasi vahelises homoloogias. Metagenoomika ja metatranskriptoomika puhul (*shotgun* sekveneerimine) järjestused assambleeritakse kontiigidesse ja seejärel kaardistatakse andmebaasistatud geenide sarnasuse alusel. Metagenoomika puhul saab DNA järjestused kaardistada referentsgenoomi või geeni külge või kasutada *de novo* geenide assambleerimist. Seejärel saab hinnata genoomide või geenide suhtelist arvukust ja järjestuste potentsiaali kasutades annoteeritud andmebaase. Metatranskriptoomi uuringutel analüüsitakse mikrobioomi aktiivseid gene. Saadud RNA järjestused kaardistatakse geenide ja geneetiliste radade külge. Metatriipkoodistamise, metagenoomika ja metatranskriptoomika tulemusi kasutatakse vastavalt mikroorganismide arvukuse, geenide ja aktiivsete geenide/organismide tuvastamiseks. (Muudetud joonis Bikel jt., 2015).

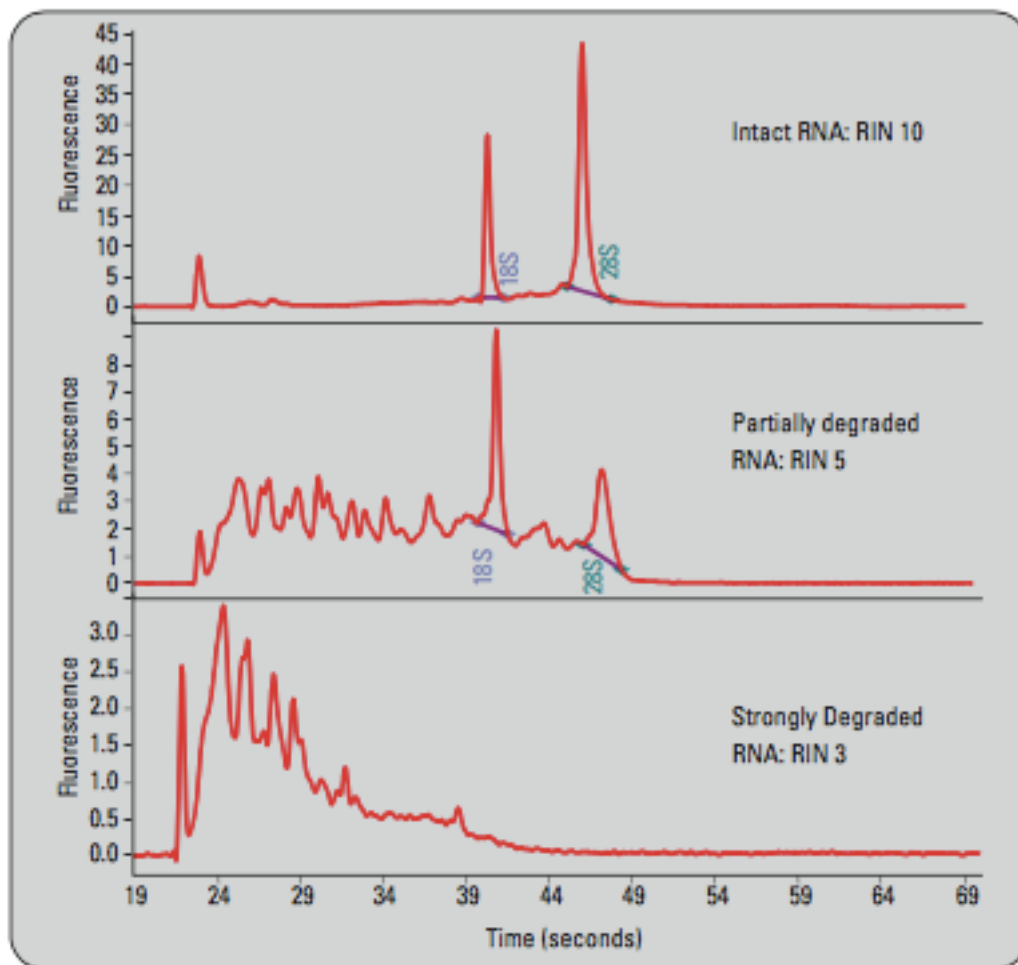
Metatranskriptoomikat kasutatakse ka inimese mikrobioomi uurimiseks. (Ottman jt., 2012). Inimese mikrobioomi väärtalitused tulenevad erinevatest haigustest. NGS meetodid pakuvad tõhusat võimalust mikrobikoosluste ja nendes toimivate muutuste uurimiseks, mistõttu võib inimese mikrobioomi kasutada kui biomarkerit haiguse varajaseks avastamiseks. (Morgan ja Huttenhower, 2012).

## 1.5 Metatranskriptoomi analüüsimeetodid ja tarkvara

### 1.5.1 RNA terviklikkus ja RIN väärtus

RNA ekspressiooni uuringud - cDNA (komplementaarne DNA) raamatukogude sekveneerimine, mikrokiiptehnoloogia ja *Northern blotting* analüüs nõuavad kõrge kvaliteediga RNA-d ja seega on kvaliteetse RNA eraldamine väga oluline. RNA analüüsimine geelelektroforeesil denatureerival agarosgeelil on kõige levinum meetod RNA terviklikkuse kontrollimiseks, mis annab tulemuseks selged 28S ja 18S rRNA transkriptide bändid. Selle meetodi puuduseks on vajamineva RNA suur kogus, mis on tavapäraselt umbes 200 ng.

Alternatiivne meetod RNA analüüsimisele geelelektroforeesil on Agilent 2100 Bioanalyzer (Agilent Technologies), mis näitab RNA kontsentratsiooni ja puhtust, mille põhjal arvutatakse RIN väärtus. Selle meetodi kasutamisel on eeliseks vaid 1-10 ng/µl proovi kohta vajaminev RNA kontsentratsioon ja andmete kujutamine nii geelelektroforeesi pildina, tabeli kujul kui ka elektroferogrammina. (<https://www.thermofisher.com/ee/en/home/references/ambion-tech-support/rna-isolation/tech-notes/is-your-rna-intact.html>). RIN väärtus ehk RNA intaktsusnumber on oluline geeniekspressiooni mõõtmise vahend. (Schroeder jt., 2006). See on meetod RNA kvaliteedi hindamiseks skaalal 1-10, kus 1 tähistab lagunenu RNA-d ja 10 intaktset ehk hea kvaliteediga RNA-d. Protsess algab RIN väärtuste valideerimisest, et saavutada läviväärtus õnnestunud ja ebaõnnestunud katsete vahel. Seda saab teha mikrokiiptehnoloogiaga või RT-PCR-ga (pöördtranskriptsiooni polümeraasi ahelreaktsioon). Peale läviväärtuse saavutamist saab seda kasutada RNA QC (RNA sekveneerimistulemuste kvaliteedikontrolli meetod) testis. Kõik RIN väärtused, mis on suuremad kui läviväärtus, läbivad QC testi. Läviväärtusest madalamad RIN väärtused eemaldatakse. (<https://www.agilent.com/cs/library/applications/5989-1165EN.pdf>).



**Joonis 2. RNA analüüs Agilent 2100 Bioanalyzer masinal.** Ülemine elektroferogramm näitab parimat võimalikku tulemust ehk mittelagunenud RNA-d väärtusega 10. Keskmine kujutab osaliselt lagunenud RNA-d RIN väärtusega 5 ning alumine tugevalt lagunenud RNA-d RIN väärtusega 3. (<https://www.agilent.com/cs/library/applications/5989-1165EN.pdf>).

Bioloogilisi RNA proove säilitatakse põhiliselt  $-20\text{ }^{\circ}\text{C}$ ,  $-80\text{ }^{\circ}\text{C}$  juures või vedelas lämmastikus (LN) (Hernandez jt., 2009). Vedelas lämmastikus proovide säilitamine hoiab proove temperatuuril, mis jääb alla vee klaasistumispunkti ehk umbes  $-135\text{ }^{\circ}\text{C}$  juures, peatades bioloogilise aktiivsuse ja takistades bioloogilise proovi lagunemist. (<https://www.thermofisher.com/blog/biobanking/is-liquid-nitrogen-the-only-gold-standard-for-long-term-sample-storage/>). Tuleks leida alternatiivseid meetodeid proovide säilitamiseks, mis oleksid stabiilsemad ja annaksid tulemuseks eraldatud RNA parima võimaliku kvaliteedinäitaja. RNA kvaliteedi ehk RIN väärtuse mõju transkriptoomi sekveneerimisele on uuritud ka enne käesolevat bakalaureusetööd. Üldiselt on teada, et RIN väärtus mõjutab edasist cDNA raamatukogude sünteesi ja saadud tulemusi. Kuidas aga mõjutavad proovide erinevad säilitusmeetodid sünteesitud RNA kvaliteeti ning sellest tulenevalt omakorda sünteesitavaid RNA raamatukogusid? Milliseid alternatiivseid meetodeid on võimalik bioloogiliste proovide

säilitamiseks kasutada? Antud bakalaureusetöös säilitati bioloogilisi proove lisaks vedelale lämmastikule, -86 °C ja -20 °C juures külmutamisele ka +4 °C juures ja TCO<sub>2</sub>-s.

### 1.5.2 Analüüsikeskkonnad ja andmebaasid

Metagenoomi ja metatranskriptoomi analüüsiks on loodud mitmed käsurea pealt töötavad programmid, mis vajavad ühe või mitme programmeerimiskeele oskust. Üheks neist on metaSPAdes (metagenoomiliste andmete konveier), mida kasutatakse sekveneerimisjärgselt toorfailide assambleerimiseks enne edasist analüüsi. (<http://bioinf.spbau.ru/en/spades3.7>; Nurk jt., 2017). MG-RAST (metagenoomika analüüsiserver) on keskkond, kuhu *fasta* või *fastq* formaadis järjestuste üles laadimine ja nende normaliseerimine võimaldab ligipääsu fülogeneetilisele ja metaboolsele andmestikule. Seoses sekveneerimishindade pideva langemise ja andmemahutude suurenemisega on sellised käepärased analüüsimeetodid nagu MG-RAST olulisel kohal. (Meyer jt., 2008; Wilke jt., 2015). MG-RAST võimaldab ka mitmete metagenoomide ja genoomide võrdlevat analüüsi. (Sun jt., 2015). Võrdluseks saab tulemused ühendada KEGG (geenide ja genoomide nomenklatuuri entsüklopeedia), eggNOG (ortoloogsete rühmade andmebaas), COG (valkude funktsiooni andmebaas) ja SEED (*de novo* annotatsioonide platvorm) andmebaasidega mitmel resolutsioonitasandil. (Thomas jt., 2012). COG andmebaasi põhiline rakendus on organismide genoomides leiduvate valkude funktsiooni tuvastamine. (Tatusov jt., 2000). Lisaks võimaldab see genoomi evolutsiooni uuringuid arendada. COG andmebaas sisaldab põhiliselt bakterite ja arhede genoomset infot. (Tatusov jt., 2001). eggNOG andmebaas sisaldab lisaks infot ka eukarüootide kohta. (Jensen jt., 2008). KEGG andmebaas ühendab teatud geenid rakus olevate geneetiliste radade või kompleksidega, põhiliselt metabolismi ja ortoloogiaga. (Kanehisa ja Goto, 2000). SEED annab võimaluse avastada ja edasi arendada *de novo* annotatsioone ja sisaldab endas ka *Subsystems* andmebaasi. (Overbeek jt., 2014). Metagenoomide analüüsimiseks saab kasutada ka IGM/M (metagenoomika andmebaas) keskkonda, mis sisaldab infot nii eukarüootide, bakterite, viiruste, plasmiidide, geenifragmentide kui ka metagenoomide ja metatranskriptomide kohta. (Chen jt., 2016)

## 2. EKSPERIMENTAALOSA

### 2.1 Töö eesmärgid

Esimese projekti eesmärgid olid järgmised:

- Vaadelda kas ja kuidas RNA erinevatest säilitusmeetoditest tulenev proovide kvaliteet peegeldab ja indikeerib geeniekspressiooni analüüsi tulemusi.
- Leida tavapärastele proovide säilitusmeetoditele võimalikke alternatiivseid meetodeid, mis parandaksid eraldatud RNA kvaliteeti.
- Identifitseerida bioloogilistes proovides leiduvad organismid ja tuvastada nende erinevaid funktsioone mullas vaadeldes analüüsikeskkondade MG-RAST ja BigTableViewerLite tõhusust metatranskriptomiliste proovide analüüsimisel.

Teise projekti eesmärgid olid:

- Identifitseerida bioloogilistes proovides leiduvad organismid ja tuvastada nende erinevaid funktsioonid mullas.

### 2.2 Materjal ja meetodika

#### 2.2.1 Proovialad ja töötused

Esimese projekti jaoks korjati Ravila tänaval asuva Tartu Ülikooli keemia instituudi vastas olevast hooldamata pargialalt mullaproovid, mis säilitati kindlate ajaintervallide tagant kuivas jääs (TCO<sub>2</sub>), -20 °C, -86 °C, 4 °C juures ja vedelas lämmastikus (Tabel 1). Proovid korjati kahelt alalt, punktist A ning punktist B, mille omavaheline kaugus oli umbes 80 meetrit (Joonis 3).



**Joonis 3. Proovivõtukohtade asukohad kaardil.** Proovid korjati kahest punktist, proovivõtukohtast A ja proovivõtukohtast B, mille omavaheline kaugus oli umbes 80 meetrit ja mis on kaardil tähistatud kahe märgisega.

**Tabel 1:** Korjatud bioloogiliste proovide säilitusviisid. Pärast näidatud aega viidi proovid  $-86^{\circ}\text{C}$  juurde kuni RNA eraldamiseni.

Säilitamise viis	1 minut	10 minutit	40 minutit	4 tundi	24 tundi	5 päeva
TCO <sub>2</sub>	+					
$-20^{\circ}\text{C}$	+	+	+	+	+	+
$-86^{\circ}\text{C}$		+				
$4^{\circ}\text{C}$		+	+	+	+	+
LN		+	+	+	+	+

Teine osa tööst oli varasemast pooleli jäänud 36 proovi metatranskriptomiline analüüs. Proovid olid korjatud ujutamiskatsetega märgalalt (kuus ujutatud ala ja kuus kontrollala) Ülo Manderi tööühma poolt Agalis kolmel eri ajal enne, nädala pärast ja mitu kuud pärast ujumiskatset ning koheselt vedelasse lämmastikku külmutatud. RNA eraldus ja muu analüüsile eelnev töö olid Dr. Ingrid Liivi poolt tehtud. cDNA raamatukogud olid valmistatud ja sekveneeritud Eesti Biokeskuses, kasutades polyA praimereid. Metatranskriptoomi analüüsi teostas in periodil 25.02-01.03.2019 Tšehhis, *The Czech Academy of Sciences* instituudis, Dr. Petr Baldriani ja Dr. Tomáš Vetrovsky juhendamisel.

## 2.2.2 Fenool-kloroform lahuse valmistamine

RNA eraldamiseks Qiagen *Rneasy PowerSoil Total RNA Kit* komplektiga valmistati eelnevalt fenool-kloroform segu, mille pH oleks vahemikus 6,5-8 ning lõppmaht 6 proovi eralduseks oleks 25 ml. Selleks võeti fenool, kloroform ja isoamüülalkohol ruumalasuhtega 25:24:1, ehk kokku segati 12,5 ml fenooli, 12 ml kloroformi ja 0,5 ml isoamüülalkoholi. Seejärel lisati 4 ml destilleeritud vett ja oodati 10 min vesilahuse tekkeni. Mõõdeti pH ja lisati vastavalt vajadusele *Tris* lahust, mille 1 ml tõstis pH-d umbes 0,2 ühiku võrra.

## 2.2.3 RNA ja DNA eraldus

RNA eraldati Qiagen *Rneasy PowerSoil Total RNA Kit* komplekti protokoll järgi. Selleks lisati kaalutud 2,0 grammi mullaproovi 15 ml graanultuubidesse. Proovidele lisati 2,5 ml *Bead Solution* lahust, 0,25 ml *SR1* lahust, 0,8 ml *IRS* lahust ja 3,5 ml fenool-kloroform-isoamüülalkohol segu. Proovidele tehti *vortex* maksimaalsel kiirusel 15 min. Seejärel tsentrifuugiti 10 min 2500 x g. Supernatant koguti uude tuubi ja lisati 1,5 ml *SR3* lahust, misjärel tehti uuesti *vortex*. Proove inkubeeriti 6 °C juures 10 min ja seejärel tsentrifuugiti 10 min 2500 x g. Lisati 5 ml *SR4* lahust ja segati ning inkubeeriti toatemperatuuril 30 min. Seejärel tsentrifuugiti 30 min 2500 x g. Supernatant eemaldati ja proove kuivatati paberil 5 min. Seejärel lisati 1 ml *SR5* lahust ja pandi proovid 10 min 45 °C vesivanni ning suspendeeriti seni kuni sade segunes. Uute tuubide sisse pandi kolonnid ja lisati 2 ml *SR5* lahust ning oodati kuni see läbi kolonnide tilgub. Lisati RNA proov, mis samuti läbi kolonni tilkus, peale mida lisati taas 1 ml *SR5* lahust ja lasti läbi tilkuda. Kolonn asetati uude tuubi ja lisati 1 ml *SR6* lahust, mis tilkus samuti läbi kolonni. Saadud RNA viidi uude 2,2 ml tuubi ja lisati 1 ml *SR4* lahust, mida segati ja inkubeeriti -20 °C juures üle öö. Järgnevalt eraldati ka DNA Qiagen *Rneasy PowerSoil DNA Elution Kit* kasutades. Selleks pandi RNA eralduses kasutatud kolonnid uutesse tuubidesse ja lisati 1 ml *SR8* lahust, mis tilkus läbi kolonni tuubi. Saadud DNA pandi 2,2 ml tuubidesse ja tsentrifuugiti 15 min 13000 x g. Supernatant eemaldati ja proove kuivatati 10 min paberil. DNA resuspendeeriti 100 µl *SR7* lahuses. Järgmisel päeval tsentrifuugiti RNA proove 15 min 13000 x g. Supernatant eemaldati ja proove kuivatati paberil 10 min. RNA resuspendeeriti 100 µl *SR7* lahuses. Kõikidele RNA proovidele tehti koheselt ka DNase töötlus. Vedelikus säilitatud proovidel kasutati lisaks ka Qiagen *LifeGuard Soil Preservation Solution* protokoll, mille kohaselt tuli proovidele lisada 5 ml *LifeGuard Soil Preservation Solution* lahust ja teha *vortex* kuni terve proov on segunenud. Seejärel proove tsentrifuugiti 5 min 2500 x g ja *LifeGuard Soil Preservation Solution* lahust eemaldati tuubist. Jätkati vastavalt Qiagen *Rneasy PowerSoil Total RNA Kit* komplekti tootjapoolsele protokollile. Antud metoodikat kasutati mõlemas uuringus.

#### 2.2.4 DNase töötlus

Proovidele tehti genoomsete DNA jääkide eemaldamiseks DNase töötlus. Proovide DNase töötluks kasutati Qiagen produkti *Dnase Max Kit*. Omavahel segati 1 µl *DNase I* ensüümi ja 10 µl *10X DNase* puhvrit. Seejärel lisati 89 µl *RNase-free* vett tuues lõppmahu 100 mikrolitrini ja proove inkubeeriti 37 °C juures 20 min. Lisati 10 µl *DNase Removal Resin* lahust ja inkubeeriti toatemperatuuril 10 min. Iga 1-2 minuti järel proove keerati ja 10 min möödudes tsentrifuugiti proove 1 min 13000 x g. Supernatant tõsteti uude tuubi, misjärel oli RNA valmis edasiseks analüüsiks.

#### 2.2.5 RNA kvaliteedikontroll

RNA kvaliteedikontrolliks saadeti proovid Eesti biokeskusesse. Teostati ScreenTape analüüs eukarüootse RNA analüüsiks, millest saadi iga proovi kvaliteedinäitaja RIN väärtus (Tabel 2).

**Tabel 2:** RNA ScreenTape analüüsi tulemused mitme väärtusega väljadel.

Meetod ja aeg	RIN väärtus (proovivõtukoht A)	RIN väärtus (proovivõtukoht B)
TCO <sub>2</sub> , 1 minut	7,3	7,2
-20 °C, 1 minut	6,8	7,6
-20 °C, 10 minutit	7,3	7,8
-20 °C, 40 minutit	7,1	6,9
-20 °C, 4 tundi	5,7	6,3
-20 °C, 24 tundi	6,1	6,9
-20 °C, 5 päeva	6,8	5,6
-86 °C, 10 minutit	7,4	7,1
4 °C, 10 minutit	7,9	6,2
4 °C, 40 minutit	7,5	7,6
4 °C, 4 tundi	7,5	5,3
4 °C, 24 tundi	7,0	7,8
4 °C, 5 päeva	7,4	6,6
LN, 10 minutit	7,5	5,3
LN, 40 minutit	7,3	7,5
LN, 4 tundi	7,5	6,7

LN, 24 tundi	8,0	2,5
LN, 5 päeva	8,3	8,1

## 2.2.6 cDNA süntees ja kontsentratsioonide mõõtmine

cDNA sünteesiks kasutati esmalt Solis BioDyne *FIREScript RT cDNA Synthesis Kit* komplekti. Kõigepealt segati kokku 3 µl RNA-d, 1 µl *random* praimereid ja 12 µl nukleasivaba vett. Protokoll lubas kasutada nii *random* kui ka *oligo(dT)* praimereid, Antud töös kasutati *random* praimereid, sest need amplifitseerivad kõiki erinevaid RNA-sid (mRNA, rRNA, *small RNAs*). *Oligo(dT)* praimerid on vaid mRNA-spetsiifilised. Proove inkubeeriti 5 min 65 °C ja seejärel asetati jääle. Lühikese tsentrifuugimise järel lisati 2 µl 10x reaktsioonipuhvrit koos *DTT* lahusega, 0,5 µl *dNTP* segu, 1 µl *FIREScript RT* lahust ja 0,5 µl *RiboGrip RNase Inhibitor* lahust. cDNA sünteesiks kasutati PCR (polümeraasi ahelreaktsioon) programmi 25 °C 10 min, 42 °C 30 min ja 85 °C 5 min. *FIREScript RT* tekitas üheaahelalise cDNA ja ei sisaldanud ensüüme, mida oleks olnud vaja teise ahela sünteesiks. Edasiseks RNA raamatukogude sünteesiks oli vaja dsDNA-d; seega tuli leida mooduseid, kuidas ssDNA-st (üheaahelalisest DNA-st) saada dsDNA (kaheaahelaline DNA) ilma spetsiaalse kitita. Vajalikud meetodid pakkus välja Dr. Karin Seegel firmast Solis BioDyne. Esialgset katsetused teostati juhuslikult valitud kahe RNA prooviga. Katsetusmeetodeid oli 6 ja seetõttu tuli cDNA-d juurde sünteesida kasutades Solis BioDyne *FireScript RT cDNA Synthesis Kit* komplekti. Seejärel mõõdeti antud proovide kontsentratsioonid kasutades Qubit® 2.0 fluoromeetrit (Thermo Fisher Scientific Inc., USA). Tulemusteks oli proov 1 puhul 13,3 ng/µL ja proov 2 puhul 11,8 ng/µL. Järgnevalt segati kokku 6x2 reaktsioonisegu erinevatel meetoditel dsDNA tegemiseks. Segu sisaldas 0,5 µl 10x *BufferB* lahust, 0,5 µl 25 mM MgCl<sub>2</sub>, 4 µl sünteesitud cDNA segu, 0,2 µl *FIREPol Pol* lahust ning 4,8 µl destilleeritud vett. Seejärel teostati katsetused ja mõõdeti PCR produktide kontsentratsioonid. Vastavad katsetusmeetodid koos tulemustega on välja toodud allolevas tabelis (Tabel 3). Qubit® masinaga mõõdetud kontsentratsioonid olid väga madalad, mis viitas sellele, et tegu oli praimeri dimeeridega ja antud meetodid ei ole efektiivsed.

**Tabel 3:** Katsetusmeetodid cDNA kaheahelaliseks muutmiseks

Meetod	1. osa	2. osa	Korduste arv	Proov number 1 kontsentratsioon (ng/ $\mu$ L).	Proov number 2 kontsentratsioon (ng/ $\mu$ L).
1.	95 °C 5 sek	25 °C 10 min	1	1,33	1,044
2.	95 °C 5 sek	25 °C 3 min	5	1,044	0,994
3.	95 °C 5 sek	25 °C 3 min	10	0,7	1,004
4.	95 °C 5 sek + jahutamine jääl	25 °C 10 min	1	0,882	0,676
5.	95 °C 5 sek + jahutamine jääl + 25 °C 5 min	95 °C 1 sek ja 25°C 3 min	5	1,184	0,698
6.	95 °C 20 sek + jahutamine jääl + 25 °C 5 min	95 °C 1 sek ja 25 °C 3 min	10	1,288	0,738

### 2.2.7 RNA raamatukogud

Kuna eelnev meetod ei töötnud kasutati RNA-st cDNA raamatukogude tegemiseks New England Biolabs *NebNext Ultra II RNA Kit* komplekti, mis tagab ka ribosomaalse RNA eemaldamise proovidest. Kasutati *random* praimereid eesmärgiga tuvastada nii prokarüoote kui eukarüoote. Varasemalt olid märgala katse cDNA raamatukogud tehtud Eesti Biokeskuses kasutades Illumina *TruSeq Stranded mRNA Kit* komplekti ja polyA praimereid, mis näitavad peamiselt eukarüootide transkripte. Protokolliga kohaselt vajasisid proovid RIN väärtusega 2-6 esimeses PCR tsüklis lühemaajalist fragmenteerimist, kui proovid, mille RIN väärtus oli 7 ja enam. Samuti tuli teha 20 ng/ $\mu$ l lahjendused (Lisa 1). Madala kontsentratsiooniga proovidele lahjendusi ei tehtud. Proovid sorteeriti RIN väärtuste alusel.

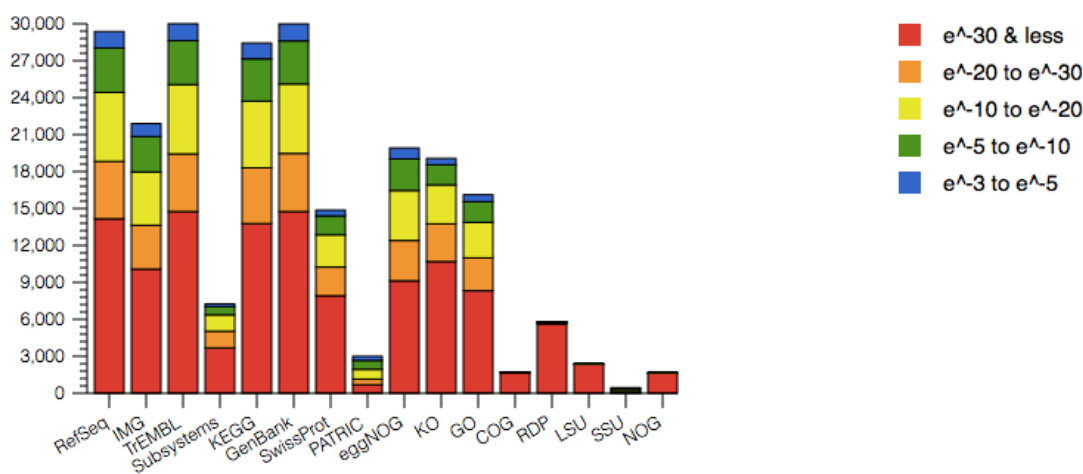
Töö teostati vastavalt ametlikule protokollile 8-16 proovi kaupa. Jääl segati kokku 5  $\mu$ l lahjendatud RNA-d, 4  $\mu$ l *NEBNext First Strand Synthesis* reaktsioonipuhvrit ja 1  $\mu$ l *random* praimereid, seejärel asetati proovid termotsüklerisse. Proovid, mille RIN väärtus oli 2-6, vajasisid 94 °C juures 7-8 min ja RIN väärtusega 7 või enam, 94 °C 15 min inkubeerimist. Seejärel pandi proovid koheselt jääle ja jätkati üheaahelalise cDNA sünteesiga. Selleks hoiti proove jääle ja lisati 8  $\mu$ l vett ja 2  $\mu$ l *NEBNext First Strand Synthesis Enzyme Mix* lahust. Peale segamist inkubeeriti

proove termotsükleris 25 °C 10 min, 42 °C 15 min, 70 °C 15 min. Seejärel jätkati koheselt kaheaheelalise cDNA sünteesiga. Selleks segati saadud üheaheelalisele produktile juurde 8 µl *NEBNext Second Strand Synthesis* reaktsioonipuhvrit, 4 µl *NEBNext Second Strand Synthesis Enzyme Mix* lahust ja 48 µl nukleasivaba vett, proove segati ja inkubeeriti termotsükleris 16°C juures 1 tund. RNA lahust tõsteti ümber 2 ml tuubidesse, lisati 144 µl *NEBNext Sample Purification Beads* lahust ja segati ning inkubeeriti 5 min toatemperatuuril. Proovid pandi magnetrestile ja oodati kuni segu selgineks; seejärel eemaldati supernatant ja lisati 200 µl 80% etanooli, inkubeeriti toatemperatuuril 30 sek ja eemaldati etanool. Etanooliga pesemise protsessi korrati. Proove kuivatati tuubid avatult 5 min ja seejärel eemaldati magnetaluselt. Lisati 53 µl 0,1X *TE* puhvrit, segati, tehti lühiajaline tsentrifuug ja inkubeeriti 2 min toatemperatuuril. Proovid pandi magnetrestile ja oodati seni, kuni segu oli selginenud. Võeti 50 µl supernatanti ja pandi taas mikrotoubidesse. Proovidele lisati 7 µl *NEBNext Ultra II End Prep* reaktsioonipuhvrit ja 3 µl *NEBNext Ultra II End Prep Enzyme Mix* lahust, tervet segu pipeteeriti üles-alla üle 10 korra, et segu korralikult seguneks. Proovid läksid termotsüklerisse inkubeerima 20 °C 30 min ja 65 °C 30 min. Sellele järgnes adapterite ligeerimine, milleks tuli teha *Adaptor Dilution Buffer* kontsentratsiooni 5-kordsed lahjendused. Seejärel lisati igale cDNA-le 2,5 µl saadud lahjendust, 1 µl *NEBNext Ligation Enhancer* lahust ja 30 µl *NEBNext Ultra II Ligation Master Mix* lahust. Pipeteeriti üles-alla üle 10 korra. Seejärel pandi proovid termotsüklerisse 20 °C juurde 15 min. Lisati 3 µl *USER Enzyme* lahust, segati ja pandi uuesti termotsüklerisse 37 °C juurde 15 min. Lisati 87 µl *NEBNext Sample Purification Beads* lahust ja segati; inkubeeriti toatemperatuuril 5 min. Proovid pandi magnetrestile ja oodati kuni segu selgineb. Seejärel segu eemaldati ja lisati 200 µl 80% etanooli; inkubeeriti 30 sek ja eemaldati etanool. Etanooliga pesemise protsessi korrati. Proove kuivatati lahtiselt 5 min ja seejärel eemaldati magnetilt. Lisati 17 µl 0,1X *TE* lahust ja segati, inkubeeriti 2 min toatemperatuuril ja asetati taas magnetile, seniks kuni proovid on selginesid. 15 µl supernatanti tõsteti uutesse PCR tuubidesse ja lisati 25 µl *NEBNext Ultra II Q5 Master Mix* lahust ja 10 µl indekseid, segati ja sellele järgnes PCR üks tsükkel 98°C 30 sek, kuus tsükli 98 °C 10 sek ja 65 °C 75 sek ning 65 °C 5 min. Proovidele lisati 45 µl *NEBNext Sample Purification Beads* lahust, suspendeeriti ja inkubeeriti toatemperatuuril 5 min. Proovid asetati magnetrestile ja kui sade segunes, eemaldati supernatant. Lisati 200 µl 80% etanooli, oodati 30 sek ja eemaldati etanool. Etanooliga pesemist korrati. Proove kuivatati tuub avatult 5 min ja seejärel lisati 23 µl 0,1X *TE* lahust ja inkubeeriti 2 min. Proovid pandi seejärel uutesse mikrotoubidesse ja säilitati -20 °C juures.

## 2.2.8 Bioinformaatiline analüüs

Esimese projektiga tekkis cDNA-de sünteesimisel eelnevalt töös lahti seletatud probleeme (vt ptk 2.2.6), millele oli raske kiiret lahendust leida ja seetõttu oli protsess aeganõudvam, kui algselt planeeritud. Analüüsi tuli alustada Ingrid Liivi poolt lõpetamata märgala katse proovide cDNA raamatukogude sekveneerimistulemuste analüüsimisest. See aga andis hea võimaluse näha, kuidas metatranskriptoomiline analüüsi käib, et hiljem enda proovide peal sama meetod läbida.

Saadud sekveneerimistulemuste analüüsimiseks MG-RAST keskkonnas on vaja toorfailid eelnevalt lahti pakkida ja seejärel ka assambleerida kasutades metaSPAdes programmi. metaSPAdes programmi kasutamisel andis häid näpunäiteid bioinformaatik Reidar Anderson. Assambleerimine on vajalik, sest üksikute järjestuste analüüsimine MG-RAST keskkonnas ei annaks soovitud tulemust. Üksiku järjestuse pikkus on keskmiselt 90 aluspaari, aga tüüpilise metagenoomse kontiigi pikkus pärast assambleerimist on keskmiselt ligi 450 aluspaari (Tabel 5). Assambleerides saab kõik failid omavahel ühendada. Järjestuse pikkusest sõltub annoteerimise kvaliteet ehk e-väärtus, mis on hea vahemikus  $1e^{-20}$  kuni  $1e^{-30}$  (Joonis 4). See pole saavutatav üksikute järjestustega, kuna need on liiga lühikesed.



**Joonis 4. Märgala katse proovide MG-RAST analüüsi tulemused.** X-teljel on erinevad andmebaasid, milles vastavaid annoteeritud lugemeid leiti. Y-teljel on lugemite arv. Heaks e-väärtuseks peetakse väärtusi, mis on joonisel kujutatud punase ja oranžina. (<https://www.mg-rast.org>).

Edasine bioinformaatiline analüüs toimus Tšehhis, Prof. Petr Baldriani ja Dr. Tomaš Vetrovsky juhendamisel. MG-RAST keskkond ei tuvasta seenetaksoneid ja seetõttu kasutati ka Tomaš Vetrovsky loodud BigTableViewerLite programmi (Vetrovsky jt., avaldamata). See võimaldab täpsemalt vaadelda näiteks metabolismi või rakuliste protsessidega seotud eukarüootide liike ja

arvukust proovides. Antud programm on veel looja poolt täiustamisel ja seega ei ole see internetist alla laetav.

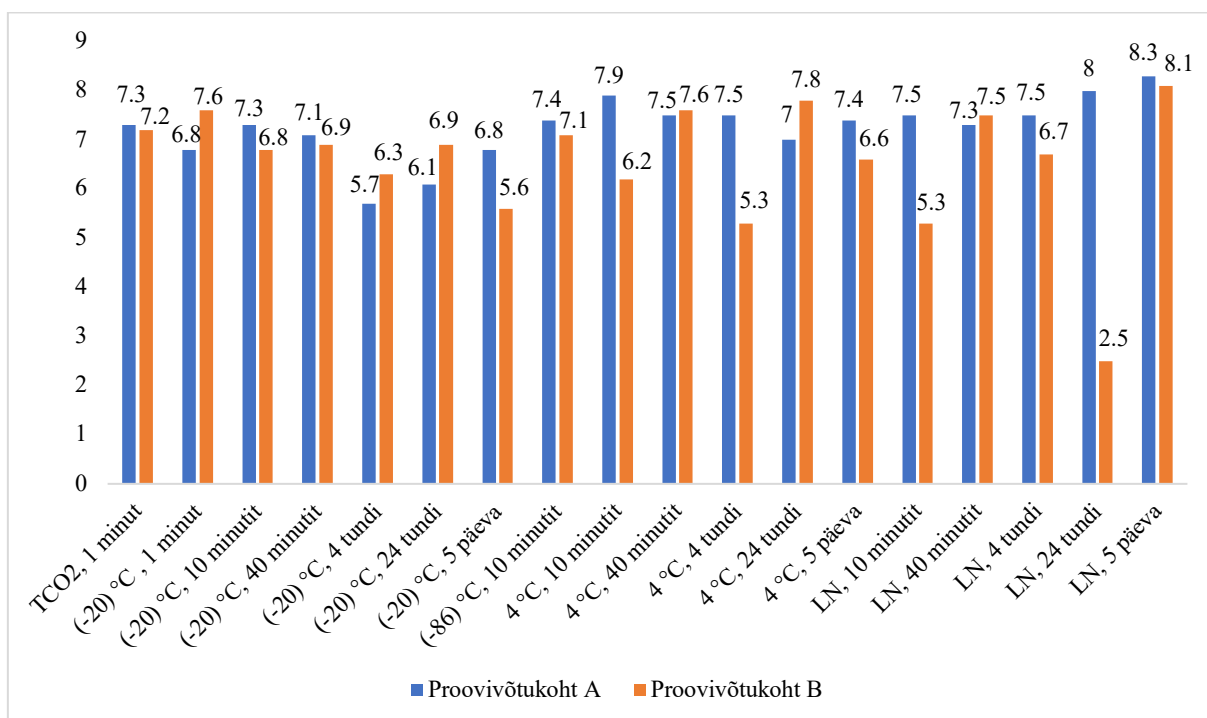
## 2.2.9 Statistiline analüüs

Statistiline analüüs viidi läbi programmpaketis Statistica (versioon 13, StatSoft Inc.). RIN väärtuste analüüsimiseks kasutasin kolmefaktorilist dispersioonanalüüsi, kus proovipunkt, aeg ja meetod olid seletavad kategoorilised tunnused. Kuna ükski faktor ei olnud statistiliselt oluline, kasutasin ka regressioonanalüüsi, et tuvastada ajafaktori mõju (aeg minutites transformeeriti log<sub>2</sub>-funktsiooniga) RIN väärtustele. Statistiliselt oluliseks pidasin olulisustõenäosust (P-väärtus) olulisusnivool  $\alpha < 0,05$ .

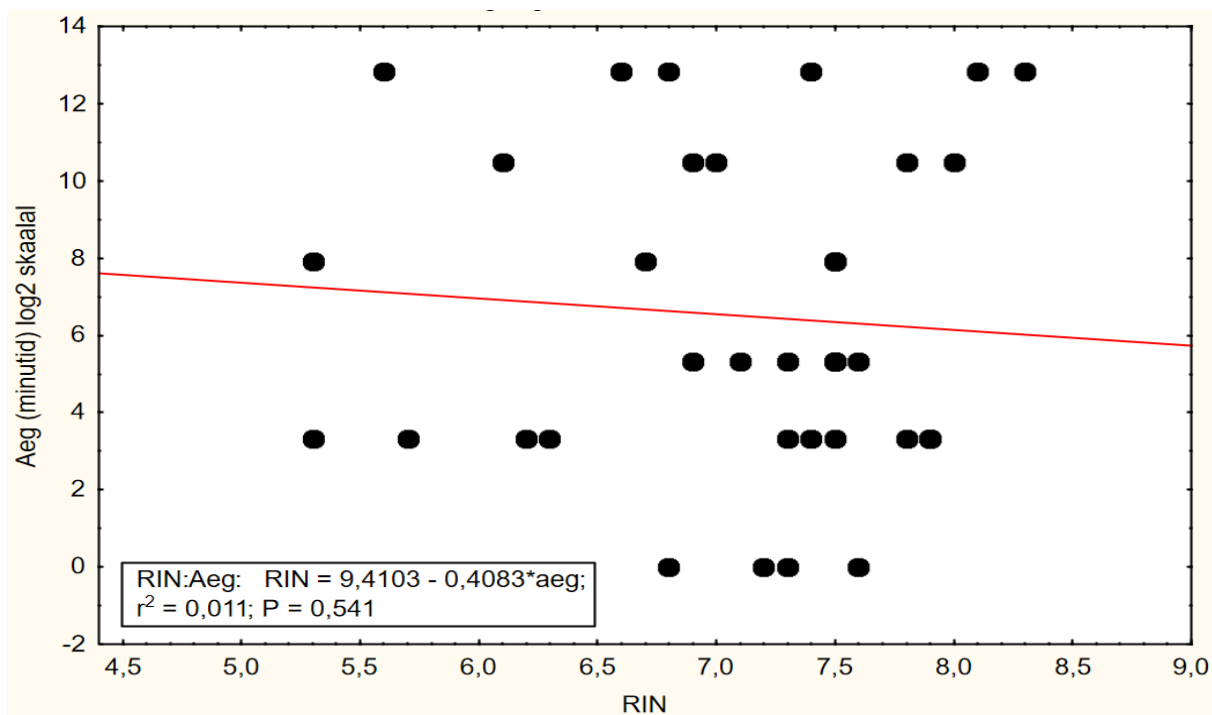
## 2.3 Tulemused ja arutelu

### 2.3.1 Esimese projekti proovide erinevate säilitusmeetodite analüüs

Antud bakalaureusetöös kasutatud bioloogilisi proove säilitati kuuel eri meetodil, millest kolm olid levinud meetodid ja 2 alternatiivset meetodit. RNA kvaliteedikontroll andis tulemuseks iga proovi RIN väärtuse ehk RNA intaktsusnumbri. (Joonis 5).



**Joonis 5. Esimese projekti proovide kvaliteet vastavalt säilitusmeetodile.** X- teljel proovide erinevad säilitusmeetodid koos ajaga, Y-teljel RIN väärtus. Sinisega on kujutatud proovivõtukoht A korjatud proovid, oranžiga proovivõtukoht B korjatud proovid. Visuaalselt on eristatav proovivõtukoht B RIN väärtuste suurem varieeruvus.



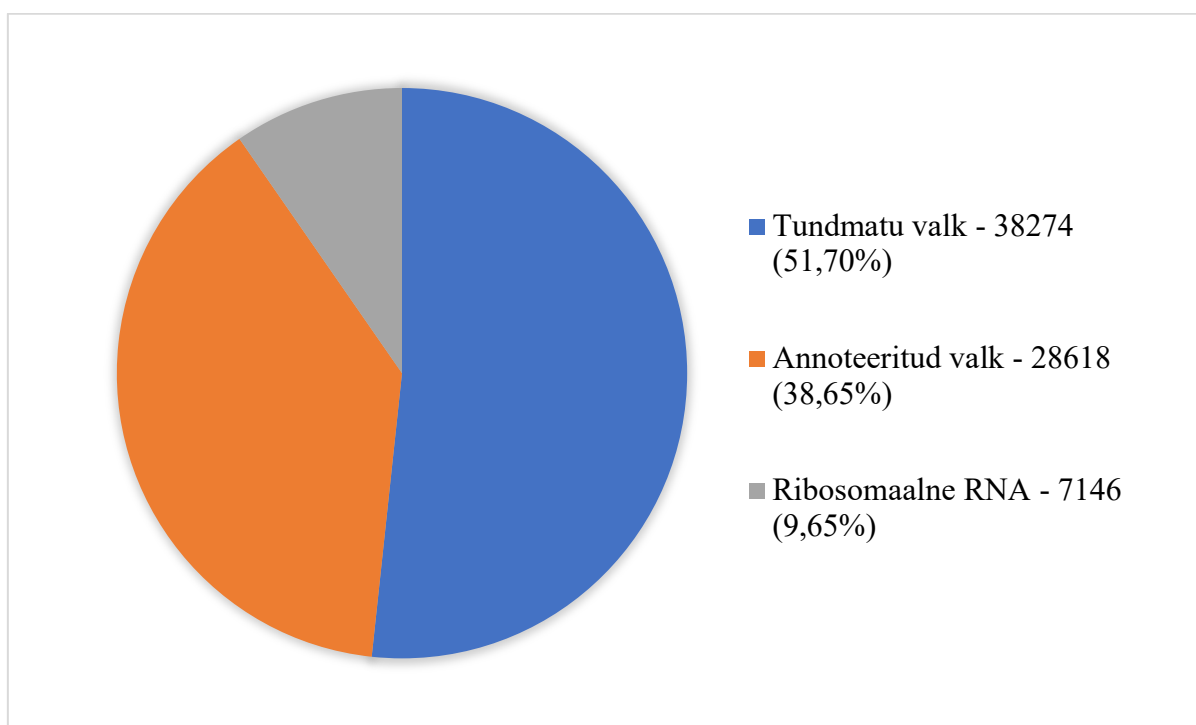
**Joonis 6.** Bioloogiliste proovide säilitamise aja mõjutatavus RIN väärtuse poolt. X-teljel on kujutatud RIN väärtused, Y-teljel aeg minutites log2 skaalal.

Antud katsete käigus taheti teada, kas proovide erinevad säilitusmeetodid mõjutavad RNA kvaliteeti RIN väärtusena ja sellest tulenevalt ka edasist cDNA raamatukogude sünteesi. Tulemuste põhjal saab öelda, et RNA kvaliteedianalüüsi tulemused ei erine märgatavalt tulenevalt erinevast säilitusmeetodist ( $F_{3,26}=0,446$ ;  $P=0,722$ ), ajast ( $F_{5,26}=0,532$ ;  $P=0,750$ ) ega erinevast proovivõtukohest ( $F_{1,26}=3,108$ ;  $P=0,090$ ). Proovivõtukohta puhul on P-väärtus küll teiste näitajatega võrreldes väiksem ja lähedasem olulisusnivoole, aga siiski mitte oluliseks peetav. Minimaalne RIN väärtus esines 24 tundi vedelas lämmastikus (proovivõtukoht B) säilitatud proovil, mille RNA on RIN väärtuse põhjal tugevalt lagunenu. See ebaloogiline erinevus võib tuleneda RNA eraldamisel tekkinud vigadest, sest RNA eraldus on väga tundlik temperatuuri ja protokollist kõrvalekaldumise poolest. Üldiselt õnnestub RNA eraldus kõigi antud bakalaureusetöös kasutatud säilitusmeetoditega. RNA-de, mis olid säilitatud  $+4^{\circ}\text{C}$  juures ja  $\text{TCO}_2$ -s tulemused ei erine suuresti tavapäraselt eelistatud säilitusmeetoditest ega ei anna silmnähtavalt paremaid tulemusi, seega alternatiivseid säilitusmeetodeid ei leitud. RNA säilitusmeetodite mõju uurimine transkriptomile on pooleli, sest andmete MG-RAST keskkonda laadimine võtab mitu nädalat.

## 2.3.2 RNA raamatukogude analüüs

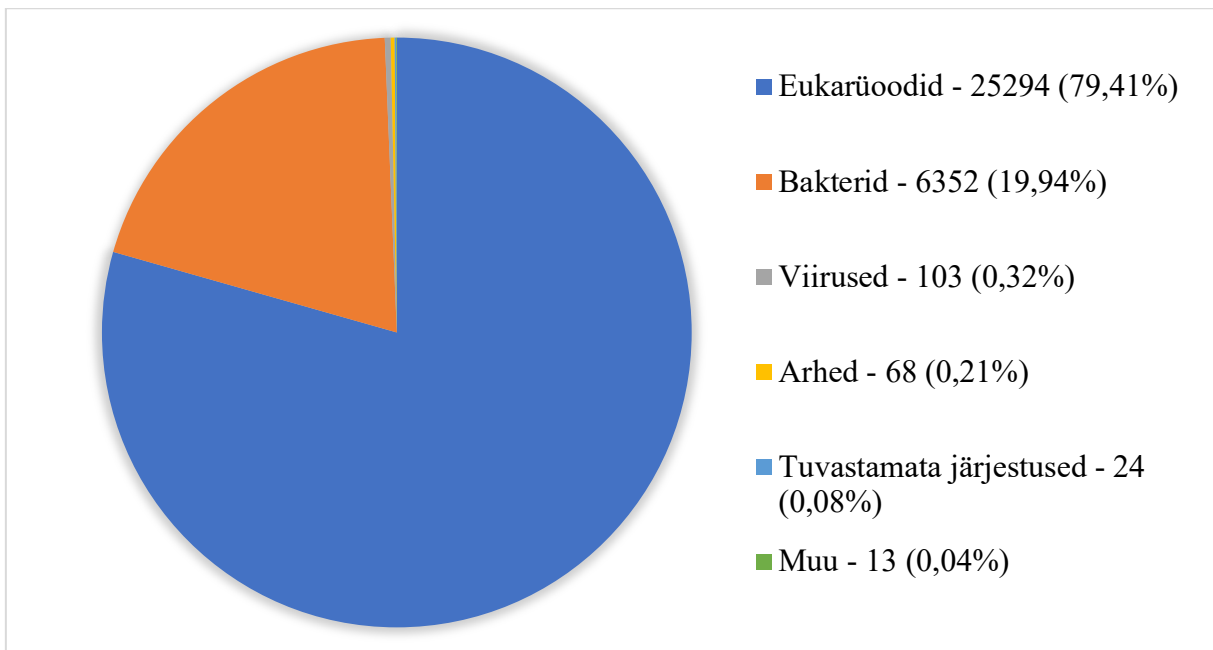
### 2.3.2.1 Märjala katse analüüsimine

MG-RAST keskkond võimaldab näha proovides säilinud ribosomaalse RNA suhtelist hulka. Metatranskriptomiliste proovide käsitlemisel on oluline, et ribosomaalne RNA saaks võimalikult suures hulgas eemaldatud. Antud proovides kogumahust oli vaid 9,65% ribosomaalne RNA, mis on väga hea tulemus ja näitab, et RNA eraldused on teostatud kvaliteetselt (Joonis 7).



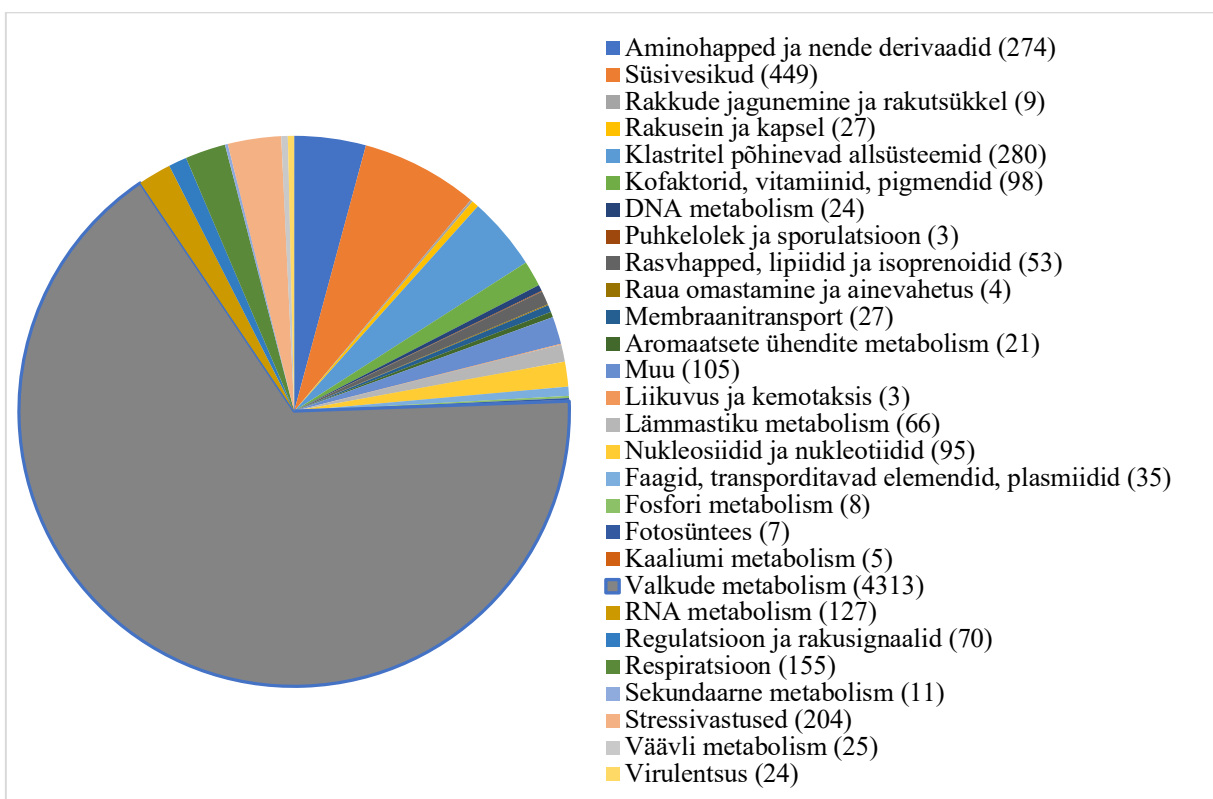
**Joonis 7.** Annoteeritud valku kodeeriva mRNA hulk märjala katse proovides. Antud proovide kogumahust oli 9,65% ribosomaalne RNA.

Järgnevalt analüüsiti MG-RAST keskkonnas proovides leiduvate eukarüootide, bakterite, viiruste ning arhede hulka (Joonis 8). Domeeni tasemel tulemused näitavad suurt eukarüootide transkriptide hulka (79,41%), lisaks ka vähesel hulgal baktereid (19,94%), viiruseid (0,32%), arhesid (0,21%) ning tuvastamata järjestusi (0,08%). Suur eukarüootide transkriptide hulk tuleneb märjala katse cDNA raamatukogude konstrueerimisel polyA praimerite kasutamisest. *Random* praimerite kasutamine, nagu seda tehti esimese projekti puhul, oleks andnud nii prokarüootide kui eukarüootide transkriptid ja oleks võimaldanud põhjalikuma analüüsi koostamist.

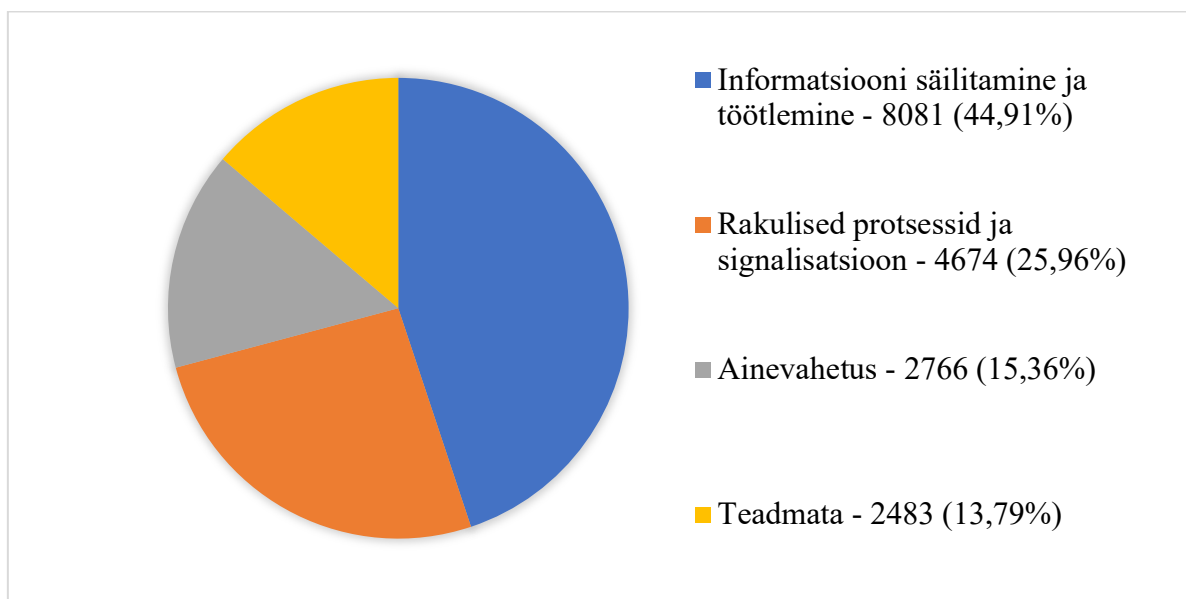


**Joonis 8.** Märgala katse tulemused domeeni tasandil. Proovid sisaldasid ülekaalukalt eukarüoote (79,41%). Eukarüootide suur arvukus proovides tulenes polyA meetodi kasutamisest.

MG-RAST paneb tulemused automaatselt vastavusse ka *Subsystems* ja COG andmebaasidega, võimaldades lisaks proovis leiduvate mikroorganismide identifitseerimisele ka erinevate funktsioonide tuvastamist. *Subsystems* tasandil analüüs näitas märgatava ülekaaluga valkude metabolismiga seotud transkripte (4313) (Joonis 9). COG analüüs näitas, et proovides leiduvad mikroorganismid on põhiliselt seotud informatsiooni säilitamise ja töötlemisega (Joonis 10).

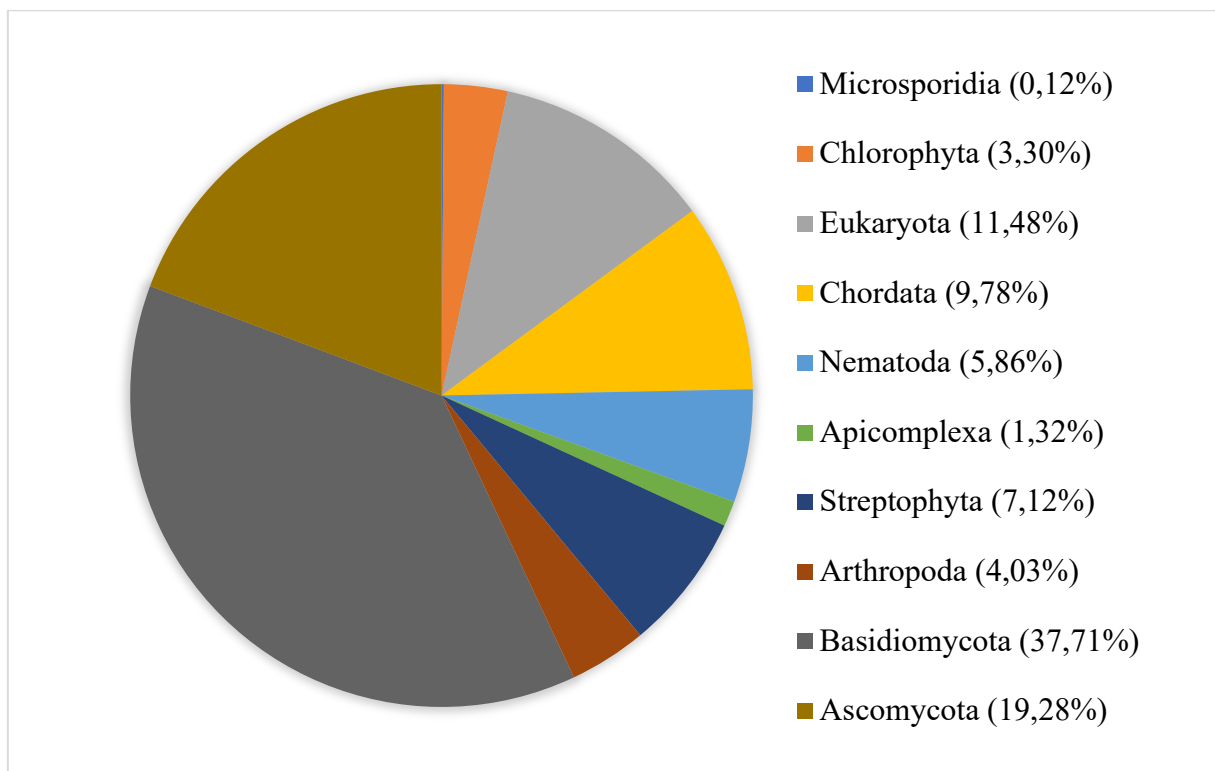


**Joonis 9.** *Subsystems* andmebaasi tulemused. Proovides leidub ülekaalukalt valkude metabolismiga seotud transkripte, lisaks ka märgataval määral mRNA-sid.

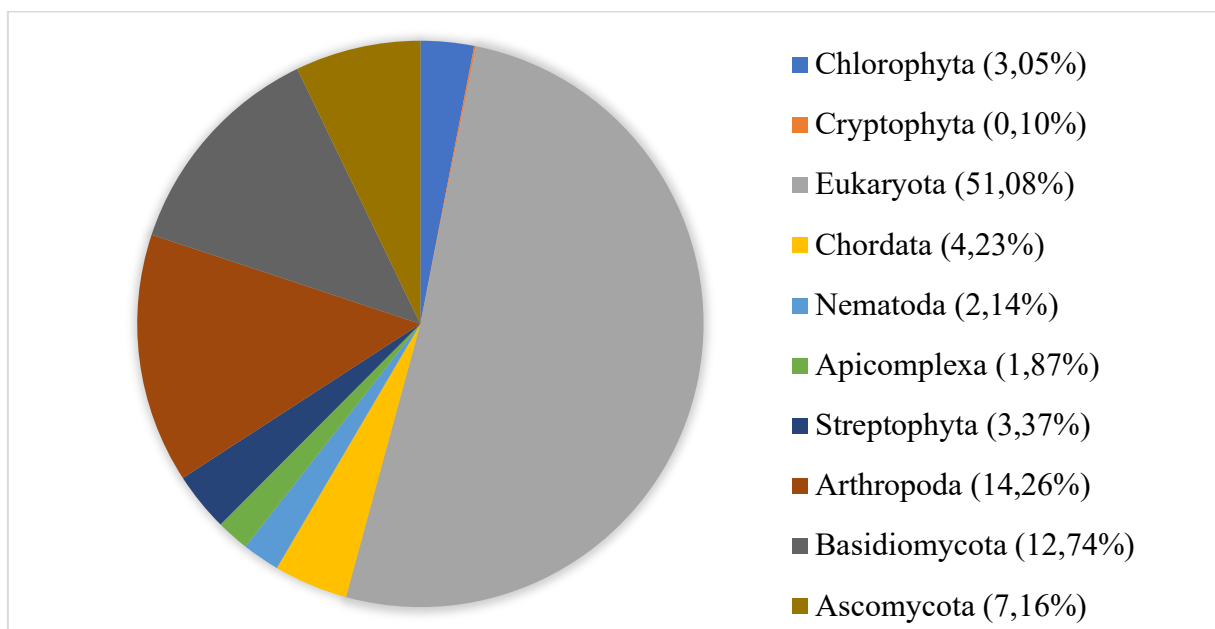


**Joonis 10.** COG analüüsi tulemused. COG analüüs näitab, et proovides leiduvad mikroorganismid on põhiliselt seotud informatsiooni säilitamise ja töötlemisega (44,91%).

Märgala proovide cDNA raamatukogud olid sünteesitud kasutades polyA praimereid, seega vaadeldi edasises analüüsis vaid eukarüoote. Esmalt uuriti antud proovides olevate eukarüootide metabolismi (Joonis 11). Metabolismiga seotud eukarüootide arvukusest domineeris *Basidiomycota* (37,71%), millele järgnesid *Ascomycota* (19,29%) ja *Eukaryota* (11,48%). Seejärel uuriti lähemalt ka rakulistes protsessides osalevaid eukarüoote. BigTableViewerLite programm näitab raku liikuvuse, kasvu, surma, transpordi, katabolismi ja rakkudevahelise suhtlusega seotud eukarüoote (Joonis 12). Rakulistes protsessides domineeris ülekaalukalt *Eukaryota* (51,08%), millele järgnesid *Arthropoda* (14,26%) ja *Basidiomycota* (12,74%).



**Joonis 11.** BigTableViewerLite tulemused metabolismiga seotud eukarüootide arvukusest proovides. Kõige rohkem on ainevahetusega seotud eukarüootidest esindatud *Basidiomycota* (37,71%) ja *Ascomycota* (19,98%).



**Joonis 12.** BigTableViewerLite tulemused rakuliste protsessidega seotud eukarüootide arvukusest proovides. Kõige rohkem on rakuliste protsessidega seotud eukarüootidest esindatud *Eukaryota* (51,08%).

Üldiselt võib näha, et MG-RAST ja BigTableViewerLite on head programmid metatranskriptomide analüüsimiseks, andes üldise ülevaate bioloogiliste proovide koostisest,

aga vajavad ka olulisel määral täiustamist. MG-RAST võimaldab näha, milliseid organisme proovides leidub ja milliste protsessidega nad seotud on, andes küllaltki pealiskaudsed tulemused. Lisaks on MG-RAST alustades failide üles laadimisest kuni analüüsi koostamiseni väga aeglane. BigTableViewerLite võimaldab proove rohkem süvitsi analüüsides identifitseerida neis leiduvaid organisme taksonoomiliselt kõrgematel tasemetel kui MG-RAST ja näha täpsemalt, millised prokarüoodid või eukarüoodid on uuritava protsessiga seotud. Samas aga on ka see programm aeglane ja analüüsi koostamine võib võtta nädalaid. Lisaks kuvab programm tulemused protsentuaalselt ning igas proovis eraldi, ehk kogu proovide analüüsi tegemiseks tuleb ise andmed kokku liita ja graafikud koostada. Mõlemad programmid vajavad palju arendamist. Ideaalis võiks selliste analüüsides koostamiseks olla programme, mis toimivad kiiremini ja võimaldavad tulemusi paremini kuvada.

### 2.3.2.2 Esimese projekti analüüsimine

Antud metatranskriptomilise analüüsi koostamisel nõudsid kõik protsessid, alates RNA eraldamisest kuni cDNA raamatukogude tegemise ja sekveneerimiseni tohutult aega. Seetõttu jõuti esimese projekti puhul ainult RIN väärtuste analüüsimiseni ja kahe esimese eesmärgi täitmiseni. cDNA raamatukogud valmistati nädala jooksul, aga sünteesitud cDNA raamatukogude sekveneerimine Eesti Biokeskuses võttis oodatust kauem aega. Sekveneerimisel kasutati HiSeq *Rapid* kiipi, millel on kaks rada ja proovid jagunevad neljaks ehk esimesel rajal on esimese ja teise lugemi järjestused L001\_R1 ja L001\_R2 ja teisel rajal esimese ja teise lugemi järjestused L002\_R1 ja L002\_R2. Antud sekveneerimine tekitab ajalises mõttes probleeme, sest ainuüksi failide allalaadimine ja lahti pakkimine võttis aega 1,5 nädalat.

### 2.3.2.3 Käimasolevad analüüsid

Poolelijäänud esimese projektiga tegeletakse eesmärkide täitmise nimel edasi. Võrreldes märgala proovide katsega peaksid esimese projekti tulemused näitama suuremat bakterite hulka, sest antud töös kasutati polyA praimerite asemel *random* primereid. See võimaldab põhjalikuma analüüsi koostamist. Nende järjestuste analüüs on hetkel faasis, kus analüüsitulemuste nägemine võtab aega veel nädalaid ja seejärel on võimalik tuvastada proovides leiduvad organismid ja nende funktsioonid mullas. Kuigi töö põhieesmärk näitas, et RIN väärtused ei sõltunud proovi säilitamise meetodist, võib oletada, et pärast mitmepäevast säilitamist on bakterite ja seente osakaal muutunud ning transkriptomis domineerivad lagunemisega seotud geenide produktid. Märgala projektis olen valmistanud samadest proovidest cDNA raamatukogud ka *random* praimeritega, mis sekveneeritakse uuema ja

võimsama Illumina NovaSeq platvormiga, et vaadelda ka peamiselt bakterite juhitud lämmastikuringet ning üleujutuste mõju sellele. HiSeq platvormi üks rada ei andnud ka piisavalt järjestusi, et valmistada geenide kogupikkusele vastavaid kontiige, mistõttu on uus sekveneerimisanalüüs vajalik ka kvaliteedi mõttes. Seejärel oleks võimalik neid kahte tööd ka põhjalikumalt omavahel võrrelda.

## KOKKUVÕTE

Metatranskriptomika on meetoodika, mis annab ülevaate mikroobikoosluse metaboolsest aktiivsusest antud ajahetkes ja kohas ning nende muutustest vastavalt keskkonnatingimustele. Metatranskriptomika võimaldab tuvastada gene, mis on ekspresseeritud, osutades ökosüsteemi tegelikule aktiivsusele. Bakalaureusetöö eesmärk oli uurida, kuidas bioloogiliste proovide erinevatest säilitusmeetoditest tulenev RNA kvaliteedinäitaja RIN mõjutab geeniekspressiooni, milliseid säilitusmeetodeid oleks võimalik kasutada suurema RIN väärtuse saavutamiseks ning tuvastada proovides leiduvad organismid ja nende funktsioonid mullas.

Bakalaureusetöö eksperimentaalses osas valmistati korjatud bioloogiliste proovide põhjal RNA raamatukogud ja jõuti järgmistele järeldustele:

- Metatranskriptomika on tõhus keskkonna geeniekspressiooni uurimiseks, aga samas ka kallis ja aeganõudev töö.
- Proovivõtt peab toimuma võimalikult kiiresti. Proovide säilitamine on keeruline, sest proovid peab kas kohe külmutama või kohe edasi töötleva.
- Bioinformaatiline analüüs nõuab tohutut arvutusressurssi.
- Proovide kvaliteet võib olla niivõrd vilets, et ei sobi edasiseks analüüsiks.
- Proovide erinevad säilitusmeetodid ei mõjuta otseselt RNA kvaliteeti RIN väärtuse põhjal.
- Proovides leiduvate organismide ja nende funktsioonide uurimiseks oleks vaja luua kiiremaid ja põhjalikumaid analüüsimeetodeid.

# Metatranscriptomic analysis of soil organisms

Keiju Rootsma

## Summary

Metatranscriptomics is a microbiological research method that gives an overview of the metabolic activity of microorganisms in certain time and place and their changes to different environmental conditions. Metatranscriptomic analysis identifies genes that are expressed referring to the actual activity of given ecosystem. The aim of this work was to analyse whether the different sample storage methods affect the RNA integrity number and the gene expression level, which type of other storage methods would give an alternative to the usual methods and also to detect the soil microorganisms and their genetically expressed activities.

The experimental part of the job focused on RNA library synthesis and the main conclusions were included:

- Metatranscriptomics is an efficient way for both prokaryotes and eukaryotes but requires a lot of time and resources.
- Sampling must be done quickly and the samples must be frozen immediately which makes the processing of samples quite difficult.
- Bioinformatical analysis requires a lot of resources.
- The extracted RNA may have very poor quality which is not suitable for metatranscriptome analysis.
- The different sample storage methods do not affect RNA integrity.
- To detect and analyse metatranscriptomic samples new analysing methods need to be developed.

## KASUTATUD KIRJANDUS

Aguiar-Pulido, V., Huang, W., Suarez-Ulloa, V., Cickovski, T., Mathee, K. and Narasimhan, G. (2016). Metagenomics, Metatranscriptomics, and Metabolomics Approaches for Mibrobiome Analysis. *Evolutionary Bioinformatics* 2016:12(S1): 5-10. doi: 10.4137/EBO.S36436.

Bailly, J., Fraissinet-Tachet, L., Verner, M-C., Debaud, J-C., Lemaire, M., Wesolowski-Louvel, M. and Marmeisse, R. (2007). Soil eukaryotic functional diversity, a metatranscriptomic approach. *The ISME Journal* (2007) 1: 632. doi: 10.1038/ismej.2007.68

Baldrian, P., Kolarik, M., Štursova, M., ... and Voriškova, J. (2012). Active and total microbial communities in forest soil are largely different and highly stratified during decomposition. *The ISME Journal* 6: 248-256 doi: 10.1038/ismej.2011.95

Bardgett, R. D., Putten, W. H. van der. (2014). Belowground biodiversity and ecosystem functioning. *Nature* 515: 505. doi:10.1038/nature13855

Bashiardes, S., Zilberman-Schapira, G. and Elinav, E. (2016). Use of Metatranscriptomics in Microbiome Research. *Bioinformatics and Biology Insights* 10: 24. doi: 10.4137/BBI.S34610.

Bikel, S., Valdez-Lara, A., Cornejo-Granados, F., Rico, K., Canizales-Quinteros, S., Soberon, X., Pozo-Yauner, L. D., Ochoa-Leyva, A. (2015). Combining metagenomics, metatranscriptomics and viromics to explore novel microbial interactions: towards a systems-level understanding of human microbiome: *Computational and Structural Biotechnology Journal* 13: 392. doi: 10.1016/j.csbj.2015.06.001

Cao, Y., Fanning, S., Proos, S., Jordan, K. and Srikumar, S. (2017). A Review on the Applications of Next Generation Sequencing Technologies as Applied to Food-Related Microbiome Studies. *Front. Microbiol.* 8: 4-5. doi: 10.3389/fmicb.2017.01829

Carey, J. (2016). News Feature: Crucial role of belowground biodiversity. *Proc Natl Acad U S A* 113: 7683. doi: 10.1073/pnas.1609238113.

Carvalhais, L. C., Dennis, P.G., Tyson, G. W., Schenk, P.M. 2013. Rhizosphere Metatranscriptomics: Challenges and Opportunities, p. 1137-1140. In Bruijn, F. J. de, Molecular Microbial Ecology of the Rhizosphere, 1st ed., vol.2.

Chen, I-M. A., Markowitz, V. M., Chu, K., ... and Kyrpides, N. C. (2016). IMG/M: integrated genome and metagenome comparative data analysis system. *Nucleic Acid Reserach* 45: D507 doi: 10.1093/nar/gkw929

Collins, F. S., Morgan, M., Patrinos, A. (2003). The Human Genome Project: Lessons from Large-Scale Biology. *Science* 300: 286-289. doi: 10.1126/science.1084564

Fierer, N., Bradford, M. A. and Jackson, R. B. (2007). Toward an ecological classification of soil bacteria. *Ecology* 88 (6): 1354-1358. <https://doi.org/10.1890/05-1839>

Geisen, S., Tveit, A. T., Clark, I. M., Richter, A., Svenning, M. M., Bonkowski, M. and Urich, T. (2015). Metatranscriptomic census of active protists in soils. *The ISME J.* **9**(10): 2178-2187. doi: 10.1038/ismej.2015.30

Handelsman, J., Rondon M. R., Brady, S. F., Clardy, J. And Goodman, R. M. (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* **5**(10): R245. [https://doi.org/10.1016/S1074-5521\(98\)90108-9](https://doi.org/10.1016/S1074-5521(98)90108-9)

Hernandez, G. E., Mondala, T. S. and Head, S. R. (2009). Assessing a novel room temperature RNA storage medium for compatibility in microarray gene expression analysis. *Biotechniques* 47 (2): 667 doi:10.2144/000113209.

Horner-Devine, M. C., Carney, K. M. and Bohannan, B. J. M. (2003). An ecological perspective on bacterial biodiversity. *Proc. R. Soc. Lond. B* (2004) **271**: 113. doi: 10.1098/rspb.2003.2549

Jensen, L. J., Julien, P., Kuhn, M., von Mering, C., Muller, J., Doerks, T. and Bork, P. (2008). eggNOG: automated construction and annotation of orthologous groups of genes. *Nucleic Acids Research* 36: D253-D254. doi:10.1093/nar/gkm796

Kanehisa, M. and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* 28 (1): 27-29. doi: 10.1093/nar/28.1.27

Kuske, C. R., Hesse, C. N., Challacombe J. F., Cullen, D., Herr, J. R., Mueller, R. C., Tsang, A., Vilgalys, R. (2014). Prospects and challenges for fungal metatranscriptomics of complex communities. *Fungal Ecology*: 133-135. <http://dx.doi.org/10.1016/j.funeco.2014.12.005>

Kutateladze, L. Y., Zakariashvili, N. G., Jobava, M. D., Burduli, T. A., Sadunishvili, T. A. (2016). Microscopic fungi spread in different types of soils in Western Georgia. *Annals of Agrarian Science* **14**: 227-229. <http://dx.doi.org/10.1016/j.aasci.2016.08.007>

Lee, C-Y., Chiu, Y-C., Wang, L-B., Kuo, Y-L., Chuang, E. Y., Lai, L-C., Tsai, M-H. (2013). Common applications of next-generation sequencing technologies in genomic research. *TCR 2* (1): 38. doi: 10.3978/j.issn.2218-676X.2013.02.09

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L. and Law, M. (2012). Comparison of Next-Generation Sequencing Systems. *J. Biomed. Sci*: 2-4. doi:10.1155/2012/251364

Manter, D. K., Delgado, J. A., Blackburn, H. D., Harmel, D., Perez de Leon, A. A. and Honeycutt, C. W. (2017). Opinion: Why we need a National Living Soil Repository. *PNAS* **114**: 13587. doi: 10.1073/pnas.1720262115

Meyer, F., Paarmann, D., Souza, M. D., ... Edwards, RA. (2008). The metagenomics RAST server- a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **2008**, **9**:2. doi:10.1186/1471-2105-9-386

Mohammadi, K., Heidari, G., Khalesro, S. and Sohrabi, Y. (2011). Soil management, microorganisms and organic matter interactions: A review. *African Journal of Biotechnology* Vol. **10**(84): 19840. doi: 10.5897/AJBX11.006

Morgan, XC., Huttenhower, C. (2012). Chapter 12: Human Microbiome Analysis. *PLoS Comput Biol* **8**(12): e1002808 doi: 10.1371/journal.pcbi.1002808

Nilsson, R. H., Anslan, S., Bahram, M., Wurzbacher, C., Baldrian, P. and Tedersoo, L. (2019). Mycobiome diversity: high-throughput sequencing and identification of fungi. *Nature Reviews Microbiology* **17**: 95-97. doi: 10.1038/s41579-018-0116-y.

Nurk, S., Meleshko, D., Korobeynikov, A. and Pevzner, P. A. (2017). metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* **27** (5): 824-825. doi: 10.1101/gr.213959.

Ottman, N., Smidt, H., de Vos, W. M. and Belzer, C. (2012). The function of our microbiota: who is out there and what do they do? *Frontiers in Cellular and Infection Microbiology*: 2. doi: 10.3389/fcimb.2012.00104.

Overbeek, R., Olson, R., Pusch, G. D., ... and Stevens, R. (2013). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology. *Nucleic Acids Research* 42: D206-D210. doi: 10.1093/nar/gkt1226

Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. *Nat. Biotechnol.* 35: 833-839. doi:10.1038/nbt.3935

Roesch, L. F. W., Fulthorpe, R. R., Riva, A., Casella, G., Hadwin, A. K., Kent, A. D., Daroub, S. H., Camargo, F. A. O., Ferriere, W. G. and Triplett, E. W. (2007). Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME J.* 1(4): 290. doi:10.1038/ismej.2007.53.

Sanger, F., Nicklen, S. and Coulson, R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74 (12): 5463-5466. doi: 10.1073/pnas.74.12.5463

Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., Lightfoot, S., Menzel, W., Granzow, M. and Ragg, T. (2006). The RIN: an RNA integrity number for assigning integrity values to measurements. *BMC Molecular Biology* 7:3: 10. doi: 10.1186/147-2199-7-3

Sun, Qinglan., Liu, L., Wu, L., Li, W., Liu, Q., Zhang, J., Liu, D., Ma, J. (2015). Web Resources for Microbial Data. *Genomics Proteomics Bioinformatics* 13: 70. doi: 10.1016/j.gpb.2015.01.008

Tatusov, R. L., Galperin, M. Y., Natale, D. A. and Koonin, E. V. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research* 28 (1): 35. doi: 10.1093/nar/28.1.33

Tatusov, R. L., Natale, D. A., Garkavtsev, I. V., Tatusova, T. A., Shankavaram, U. T., Rao, B. S., Kiryutin, B., Galperin, M. Y., Fedorova, N. D. and Koonin, E. V. (2001). The COG

database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Research* 29 (1): 22. doi: 10.1093/nar/29.1.22

Tedersoo, L., Anslan, S., Bahram, M., ... Abarenkov, K. (2015). Shotgun metagenomes and multiple primer pair-barcode combinations of amplicons reveal biases in metabarcoding analyses of fungi. *MycoKeys* 10: 1-20. <https://doi.org/10.3897/mycokeys.10.4852>

Thomas, T., Gilbert, J. and Meyer, F. (2012). Metagenomics- a guide from sampling to data analysis. *Microbial Informatics and Experimentation* 2: 7. <https://doi.org/10.1186/2042-5783-2-3>

Torsvik, V., Ovreas, L., Thingstad, T. F. (2002). Prokaryotic Diversity- Magnitude, Dynamics, and Controlling Factors. *Science* 296 (5570): 1064-1066. doi: 10.1126/science.1071698

Toseland, A., Moxon, S., Mock, T. and Moulton, V. (2014). Metatranscriptomes from diverse microbial communities: assessment of data reduction techniques for rigorous annotation. *BMC Genomics* 15: 1. doi: 10.1186/1471-2164-15-901.

Delgado-Baquerizo, M., Oliverio, A. M., Brewer, T. E., Benavent-Gonzalez, A., Eldridge, D. J., Bardgett, R. D., Maestre, F. T., Singh, B.K., Fierer, N. (2018). A global atlas of the dominant bacteria found in soil. *Science* 359: 320. doi: 10.1126/science.aap9516.

Žifčáková, L., Vetrovský, T., Howe, A. and Baldrian, P. (2016). Microbial activity in forest soil reflects the changes in ecosystem properties between summer and winter. *Environmental Microbiology* 18(1): 292-294. doi: 10.1111/1462-2920.13026.

## **KASUTATUD VEEBIAADDRESSID**

Euroopa Keskkonnaagentuur. Kasutatud 27.04.2019.

<https://www.eea.europa.eu/et/themes/soil/intro>

ThermoFisher Scientific. Is Your RNA Intact? Methods to Check RNA Integrity. Kasutatud 16.05.2019

<https://www.thermofisher.com/ee/en/home/references/ambion-tech-support/rna-isolation/tech-notes/is-your-rna-intact.html>

CD Genomics. Overview of Metatranscriptomic Sequencing: Principles, Workflow, and Applications. Kasutatud 20.05.2019

<https://www.cd-genomics.com/the-principles-workflow-and-applications-of-metatranscriptomic-sequencing.html>

ThermoFisher Scientific. Is Liquid Nitrogen The Only Gold Standard For Long-Term Sample Storage? Kasutatud 20.05.2019

<https://www.thermofisher.com/blog/biobanking/is-liquid-nitrogen-the-only-gold-standard-for-long-term-sample-storage/>

Agilent Technologies. RNA Integrity Number (RIN)- Standardization of RNA Quality Control. Kasutatud 03.05.2019

<https://www.agilent.com/cs/library/applications/5989-1165EN.pdf>

metaSPAdes

<http://bioinf.spbau.ru/en/spades3.7>

MG-RAST

<https://www.mg-rast.org>

# LISA 1

**Tabel 4:** RNA proovide 20 ng/ul lahjendused. Proovid on sorteeritud RIN väärtuste järgi.

Proovi number	RIN väärtus	RNA kontsentratsioon (ng/μl)	Lisatav RNA lahuse maht	Lisatav vee maht
12	8,3	14,4	10	0
8	8,1	44,2	10	12,1
9	8,0	3,21	10	0
14	7,9	62,0	10	21
17	7,9	59,8	10	19,9
36	7,9	56,0	10	18
16	7,8	63,7	10	21,85
23	7,8	84,1	10	32,05
6	7,6	85,0	10	32,5
33	7,6	84,2	10	32,1
7	7,5	3,97	10	0
30	7,5	117	10	48,5
31	7,5	131	10	55,5
38	7,5	44,4	10	12,2
39	7,5	48,3	10	14,15
13	7,4	126	10	53
22	7,4	61,5	10	20,75
42	7,4	102	10	41
3	7,3	47,4	10	13,7
15	7,3	68,8	10	24,4
37	7,3	73,7	10	26,85
1	7,2	71,9	10	25,95
2	7,2	88,0	10	34
32	7,1	53,2	10	16,6
41	7,1	112	10	46
29	7,0	107	10	43,5
20	6,9	88,9	10	34,35
34	6,9	47,3	10	13,65
5	6,8	122	10	51
19	6,8	124	10	52

10	6,7	6,69	10	0
35	6,7	159	10	69,5
18	6,2	54,8	10	17,4
27	6,1	80,0	10	30
26	5,7	20,0	10	0
21	5,6	64,3	10	22,15
25	5,3	133	10	56,5
40	5,3	84,1	10	32,05
4	5,1	130	10	55
28	3,2	144	10	62
24	2,4	141	10	60,5
11	0	2,75	10	0

## **Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks**

Mina, Keiju Rootsma

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose **Mullaorganismide metatranskriptoomi analüüs**, mille juhendaja on Leho Tedersoo,

reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.

2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 3.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

*Keiju Rootsma*  
**27.05.2019**