Tartu Ülikool
Loodus- ja täppisteaduste valdkond
Matemaatika ja statistika instituut

Mikk Tomson

Vasikate seedetrakti mikrobioomi mitmekesisus ja areng

matemaatiline statistika

Bakalaureusetöö (9 EAP)

Juhendaja: prof. Tanel Kaart

Tartu 2023

VASIKATE SEEDETRAKTI MIKROBIOOMI MITMEKESISUS JA ARENG

Bakalaureusetöö

Mikk Tomson

Lühikokkuvõte

Käesoleva bakalaureusetöö eesmärk on uurida lehmvasikate mikrobioomi mitmekesisuse muutumist sünnist kuni kolmanda elukuuni võrreldes seda nende emade mikrobioomi mitmekesisusega ühe piimafarmi näitel. Kõigepealt tutvustatakse kasutatavaid statistilisi analüüsimeetodeid, millele järgneb mikrobioomi mitmekesisuse analüüs. Analüüs hõlmab erinevate α- ja β-mitmekesisuse mõõdikute leidmist ning peakomponentanalüüsi läbiviimist.

CERCS teaduseriala: P160 Statistika, operatsioonianalüüs, programmeerimine, finants- ja kindlustusmatemaatika.

Märksõnad: mikrobioom, peakomponentanalüüs, lehmvasikad.

DIVERSITY AND DEVELOPMENT OF THE CALVES' INTESTINAL TRACT MICROBIOME

Bachelor thesis

Mikk Tomson

Abstract

The aim of this bachelor thesis is to study the development of the microbiome diversity of calves from birth to the third month of life, comparing it with the microbiome diversity of their mothers from the example of one dairy farm. First, statistical analysis methods are introduced, followed by microbiome diversity analysis. The analysis includes finding various α- and β-diversity measures and conducting principal component analysis.

CERCS research specialisation: P160 Statistics, operations research, programming, financial and actuarial mathematics.
Key Words: microbiome, principal component analysis, calves.

Sisukord

Sissejuhatus

Veise soolestikus on väga palju erinevaid mikroorganisme ehk mikroobe. Mikrobioomiks nimetatakse selliste mikroobide kogumit. Praeguste teaduslikke uuringute järgi ei ole lõplikult kindlaks tehtud, kas vasika soolestiku mikrobioomi arenemine algab looteeas või alles pärast sündi (Zhu *et al.*, 2021). Küll on aga teada, et soolestiku mikrobioom mängib olulist rolli vastsündinud veise immuunsüsteemi ja seedimise väljakujunemises. Samuti on leitud seoseid vasikate soolestiku mikrobioomi ja mitmete haiguste vahel (Slanzon *et al.*, 2022). Seetõttu on oluline mõista, millest oleneb mikrobioomi areng.

Käesoleva bakalaureusetöö eesmärk on uurida erinevaid võimalusi mikrobioomi mitmekesisuse statistilisel analüüsimisel. Töö esimeses osas antakse ülevaade statistilisest metodoloogiast, mida kasutatakse andmete analüüsimiseks. Töö teises pooles analüüsitakse ühe Tartumaa suure piimafarmi vasikate mikrobioomi andmeid.

Autor tänab bakalaureusetöö juhendajat Tanel Kaarti rohkete nõuannete eest.

1 Mikrobioomi mitmekesisuse analüüsimise meetodid

1.1 α-mitmekesisus

α-mitmekesisuse (ingl α-*diversity*) all mõistetakse ühe konkreetse proovi tulemuste mitmekesisust. See tähendab kui palju erinevaid unikaalseid mikroobe leidub looma soolestikus ja samuti, kuidas nad on jaotunud. Antud töös vaadeldakse α-mitmekesisuse juures kolme näitajat:

- liigirikkust;
- Shannoni entroopiat;
- Simpsoni mitmekesisuse indeksit.

Liigirikkuse all mõistetakse unikaalsete mikroobide arvu. Seejuures tuleb ära märkida, et haruldasemaid liike sekveneerimisel ei pruugita avastada ning nii on andmestikus vaadeldud liigirikkus alamhinnang tegelikule liigirikkusele. Lisaks unikaalsete liikide arvule uuritakse käesolevas töös ka nende jaotuvust. Jaotuvuse uurimise eesmärk on mõista, kas looma soolestiku mikrobioomis on kõik liigid võrdselt esindatud või on mõne liigi arvukus oluliselt suurem kui teiste. Selleks kasutakse nii Shannoni entroopiat kui ka Simpsoni mitmekesisuse indeksit. Shannoni entroopia arvutatakse kasutades valemit:

$$H = -\sum_{i=1}^{N} p_i(\log_2 p_i)$$

kus N on unikaalsete mikroobiliikide arv ning p_i on i-nda liigi osakaal proovis. Osakaal on defineeritud järgnevalt:

$$p_i = \frac{n_i}{\sum_{i=1}^{N}(n_i)}$$

kus N on unikaalsete liikide arv ning n_i on i-nda liigi kogus vastavas proovis. Osakaalud saavad olla vahemikus nullist üheni. Teatavasti pole $\log_2 0$ määratud, seetõttu on eraldi defineeritud, et Shannoni entroopia korral $0 \cdot \log_2 0 = 0$. Shannoni entroopia saavutab väärtusi nullist lõpmatuseni, kus kõrgem väärtus viitab mitmekesisemale mikrobioomile.

Simpsoni mitmekesisuse indeksi arvutamiseks kasutakse valemit:

$$D = 1 - \sum_{i=1}^{N} p_i^2$$

kus taaskord N on unikaalsete liikide arv ning p_i on i-nda liigi osakaal proovis. Simpsoni mitmekesisuse indeks saavutab väärtusi nullist üheni, kus kõrgem väärtus viitab mitmekesisemale mikrobioomile (Oksanen, 2022). Autoril ei ole põhjust arvata, et üks α-mitmekesisuse mõõdik oleks parem kui teine. Pigem täiendavad nad üksteist. Liigirikkus võtab arvesse vaid unikaalsete liikide arvu ja ei arvesta nende jaotuvust. Shannoni entroopia ja Simpsoni mitmekesisuse indeks arvestavad nii liigirikkuse kui liikide arvu jaotuvusega, kuid Simpsoni mitmekesisuse indeks paneb suuremat rõhku proovis domineerivatele liikidele.

Illustreerimiseks tuuakse järgnevalt α-mitmekesisuse mõõdikute käitumise näide. Olgu kolm erinevat proovi, kus on mõõdetud kolme erineva liigi arvukust. Proovis A esineb kõiki kolme liiki võrdselt, see tähendab osakaalud on vastavalt $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. Proovis B esineb kõiki liike erineval määral, osakaalud on vastavalt $(\frac{4}{7}, \frac{2}{7}, \frac{1}{7})$. Proovis C esineb esimest kahte liiki võrdselt, kuid kolmandat liiki ei leidu üldse, seega osakaalud on $(\frac{1}{2}, \frac{1}{2}, 0)$. Selliste hüpoteetiliste proovide α-mitmekesisuse mõõdikute tulemused on arvutatud tabelis 1.

Tabel 1: α-mitmekesisuse mõõdikute näide

	Liigirikkus	Shannoni entroopia	Simpsoni indeks
Proov A	3	$-(\frac{1}{3}(\log_2 \frac{1}{3}) + \frac{1}{3}(\log_2 \frac{1}{3}) + \frac{1}{3}(\log_2 \frac{1}{3})) \approx 1{,}58$	$1 - (\frac{1}{3})^2 + (\frac{1}{3})^2 + (\frac{1}{3})^2 = \frac{2}{3}$
Proov B	3	$-(\frac{4}{7}(\log_2 \frac{4}{7}) + \frac{2}{7}(\log_2 \frac{2}{7}) + \frac{1}{7}(\log_2 \frac{1}{7})) \approx 1{,}37$	$1 - (\frac{4}{7})^2 + (\frac{2}{7})^2 + (\frac{1}{7})^2 = \frac{28}{49}$
Proov C	2	$-(\frac{1}{2}(\log_2 \frac{1}{2}) + \frac{1}{2}(\log_2 \frac{1}{2})) = 1$	1-$(\frac{1}{2})^2 + (\frac{1}{2})^2 = \frac{1}{2}$

Siit järeldub, et nii Shannoni entroopia kui ka Simpsoni mitmekesisuse indeksi järgi on kõige mitmekesisem proovidest proov A, sellele järgneb proov B ja viimaks proov C.

1.2 β-mitmekesisus

β-mitmekesisust (β-*diversity*) kasutatakse juhtudel, kui soovitakse mõõta kahe mikrobioomi vahelist erinevust. Erinevuse mõõtmisel võetakse arvesse nii liigilist erinevust kui ka liikide jaotuste erinevust. Kõige lihtsam viis β-mitmekesisuse mõõtmiseks on proovide vahelise Eukleidilise kauguse mõõtmine. Proovide A ja B vaheline Eukleidiline kaugus arvutatakse kasutades valemit:

$$d_E(A,B) = \sqrt{\sum_{i=1}^{N}\Big(n_i(A) - n_i(B)\Big)^2}$$

kus N on unikaalsete liikide koguarv mõlema proovi peale kokku ning n_i on i-nda liigi kogus vastavas proovis.

Proovide vahelise β-mitmekesisuse mõõtmisel Eukleidilise kauguse abil võib tekkida ootamatu olukord, kus kaks proovi, millel ei ole ühtegi ühist liiki, on üksteisega sarnasemad kui proovid, kus kõik liigid on identsed. Illustreerimiseks võib ette kujutada olukorda, kus soovitakse võrrelda kolme proovi. Proovide liigilised kogused on toodud tabelis 2 ja proovide vahelised Eukleidilised kaugused tabelis 3.

Tabel 2: Proovide liigilised kogused

	liik 1	liik 2	liik 3
Proov A	0	1	0
Proov B	1	0	1
Proov C	10	0	10

Tabel 3: Eukleidilised kaugused

	Proov A	Proov B	Proov C
Proov A	0	1.73	14.18
Proov B	1.73	0	12.73
Proov C	14.18	12.73	0

Proovidest B ja C on leitud täpselt samad liigid, kuid nende omavaheline Eukleidiline kaugus on palju suurem kui proovide A ja B vahel, ehkki proovides A ja B ei leidu ühtegi ühist liiki. Tekkinud kummalist olukorda nimetatakse Orloci para-

doksiks ja see on peamiseks põhjuseks, miks ei ole soovitatav kasutada Eukleidilist kaugust β-mitmekesisuse mõõtmiseks (Orloci, 1978). Selle asemel tuleks kasutada Bray-Curtise eripära. Proovide A ja B vaheline Bray-Curtise eripära leitakse kasutades valemit:

$$d_{BC}(A,B) = 1 - \frac{2C_{AB}}{S_A + S_B}$$

kus

$$C_{AB} = \sum_{i=1}^{N} \Big(\min\big(n_i(A), n_i(B)\big) \Big)$$
$$S_A = \sum_{i=1}^{N} (n_i(A))$$
$$S_B = \sum_{i=1}^{N} (n_i(B))$$

seega

$$d_{BC}(A,B) = 1 - \frac{2\sum_{i=1}^{N} \Big(\min\big(n_i(A), n_i(B)\big) \Big)}{\sum_{i=1}^{N}(n_i(A)) + \sum_{i=1}^{N}(n_i(B))}$$

kus N on unikaalsete liikide koguarv mõlema proovi peale kokku ning n_i on i-nda liigi kogus vastavas proovis. Mõõdik omandab väärtusi vahemikus nullist üheni. Väärtus null tähendab, et proovid on identsed ning väärtus üks tähendab, et proovides ei ole ühtegi ühist liiki (Kers ja Saccenti, 2021). Mõõdikut saab leida ka kasutades koguste (n_i) asemel osakaale (p_i). Sellisel juhul

$$S_A + S_B = \sum_{i=1}^{N} (p_i(A)) + \sum_{i=1}^{N} (p_i(B)) = 1 + 1 = 2$$
$$d_{BC}(A,B) = 1 - \frac{2C_{AB}}{2} = 1 - C_{AB}$$

Illustreerimiseks tuuakse järgnevalt näide Bray-Curtise eripära arvutamise kohta. Olgu kolm erinevat proovi, kus on mõõdetud kolme erineva liigi arvukust. Proovis A esineb kõiki kolme liiki võrdselt, see tähendab osakaalud on vastavalt $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. Proovis B esineb kõiki liike erineval määral, osakaalud on vastavalt $(\frac{4}{7}, \frac{2}{7}, \frac{1}{7})$. Proo-

vis C esineb esimest kahte liiki võrdselt, kuid kolmandat liiki ei leidu üldse, seega osakaalud on $(\frac{1}{2}, \frac{1}{2}, 0)$. Antud hüpoteetiliste proovide omavahelised Bray-Curtise eripärad on arvutatud tabelis 4.

Tabel 4: Bray-Curtise eripära näide

	Proov A	Proov B	Proov C
Proov A	$1 - \frac{2(\frac{1}{3}+\frac{1}{3}+\frac{1}{3})}{1+1} = 0$	$1 - \frac{2(\frac{1}{3}+\frac{2}{7}+\frac{1}{7})}{1+1} = \frac{5}{21}$	$1 - \frac{2(\frac{1}{3}+\frac{1}{3})}{1+1} = \frac{1}{3}$
Proov B	$1 - \frac{2(\frac{1}{3}+\frac{2}{7}+\frac{1}{7})}{1+1} = \frac{5}{21}$	$1 - \frac{2(\frac{4}{7}+\frac{2}{7}+\frac{1}{7})}{1+1} = 0$	$1 - \frac{2(\frac{1}{2}+\frac{2}{7})}{1+1} = \frac{3}{14}$
Proov C	$1 - \frac{2(\frac{1}{3}+\frac{1}{3})}{1+1} = \frac{1}{3}$	$1 - \frac{2(\frac{1}{2}+\frac{2}{7})}{1+1} = \frac{3}{14}$	$1 - \frac{2(\frac{1}{2}+\frac{1}{2})}{1+1} = 0$

Oodatult on proovi võrdlemisel iseendaga Bray-Curtise eripära väärtus 0. Kõige enam erinevad üksteisest proovid A ja C.

1.3 Peakomponentanalüüs

Peakomponentanalüüsi kasutatakse olukordades, kus uuritavaid tunnuseid on väga palju ja leidub tunnuseid, mis on omavahel tugevalt korreleeritud. Meetodi eesmärk on vähendada tunnuste arvu, leides tunnuste lineaarkombinatsioone, mis kirjeldaksid võimalikult hästi algsetes tunnustes sisalduvat informatsiooni. Selliseid lineaarkombinatsioone nimetatakse peakomponentideks (Tiit ja Viil, 1992). Peakomponendid moodustakse nii, et esimene lineaarkombinatsioon kirjeldab ära võimalikult suure osa lähtetunnuste hajuvusest, teine peakomponent kirjeldab ära võimalikult suure osa alles jäänud lähtetunnuste hajuvusest ja nii edasi kuni kogu lähtetunnuste hajuvus on kaetud. Kõik peakomponendid on omavahel mittekorreleeritud. Peakomponentide leidmiseks leitakse andmetest kõigepealt kovariatsioonimaatriks. Esimese peakomponendi kordajate vektor on kovariatsioonimaatriksi suurimale omaväärtusele vastav omavektor. Teise peakomponendi kordajate vektor on kovariatsioonimaatriksi suuruselt teisele omaväärtusele vastav omavektor. Analoogselt on võimalik leida kõik peakomponendid. I-s peakomponent kirjeldab

$\frac{\lambda_i}{\sum_{j=1}^{k}\lambda_j}$ kogu hajuvusest, kus λ on kovariatsioonimaatriksi omaväärtus. Juhul kui tunnused ei ole mõõdetud samal skaalal ja/või ühikutes, siis tuleb kovariatsioonimaatriksi asemel kasutada korrelatsioonimaatriksit. (Johnson ja Wichern, 2007). Illustreerimiseks tuuakse järgnevalt näide peakomponentide leidmisest kolme tunnuse korral. Olgu uuritavate tunnuste vahelised korrelatsioonid toodud tabelis 5. Peakomponentide leidmiseks tuleb esmalt leida andmetele vastava korrelatsiooni-

Tabel 5: Tunnuste vaheline korrelatsioonimaatriks

	Liik 1	Liik 2	Liik 3
Liik 1	1	$\frac{4}{5}$	$\frac{2}{5}$
Liik 2	$\frac{4}{5}$	1	$\frac{3}{5}$
Liik 3	$\frac{2}{5}$	$\frac{3}{5}$	1

maatriksi omaväärtused.

$$\begin{vmatrix} 1-\lambda & 0,8 & 0,4 \\ 0,8 & 1-\lambda & 0,6 \\ 0,4 & 0,6 & 1-\lambda \end{vmatrix} = 0 \implies$$

$$\implies (1-\lambda)^3 + \frac{192}{1000} + \frac{192}{1000} - \frac{16}{100} + \frac{16\lambda}{100} - \frac{36}{100} + \frac{36\lambda}{100} - \frac{64}{100} + \frac{64\lambda}{100} = 0 =$$

$$= -\lambda^3 + 3\lambda^2 - 1.84\lambda + 0.224 = \begin{cases} \lambda_1 = 2,2 \\ \lambda_2 = 0,62 \\ \lambda_3 = 0,16 \end{cases}$$

Seejärel tuleb leida omaväärtustele vastavad omavektorid. Suurimale omaväärtusele vastav omavektor leitakse järgnevalt:

$$\overrightarrow{v_1} = \begin{cases} (1-2,2)x_1 + 0,8x_2 + 0,4x_3 = 0 \\ 0,8x_1 + (1-2,2)x_2 + 0,6x_3 = 0 \\ 0,4x_1 + 0,6x_2 + (1-2,2)x_3 = 0 \end{cases}$$

$$\overrightarrow{v_1} = (1,15; 1,25; 1)^T$$

Analoogselt leitakse ka teistele omaväärtustele vastavad omavektorid:

$$\overrightarrow{v_2} = (-0,66; -0,19; 1)^T$$

$$\overrightarrow{v_3} = (2,36; -2,97; 1)^T$$

Seega peakomponendid on:

$$y_1 = 1,15 \cdot \text{liik } 1 + 1,25 \cdot \text{liik } 2 + 1 \cdot \text{liik } 3$$

$$y_2 = -0,66 \cdot \text{liik } 1 - 0,19 \cdot \text{liik } 2 + 1 \cdot \text{liik } 3$$

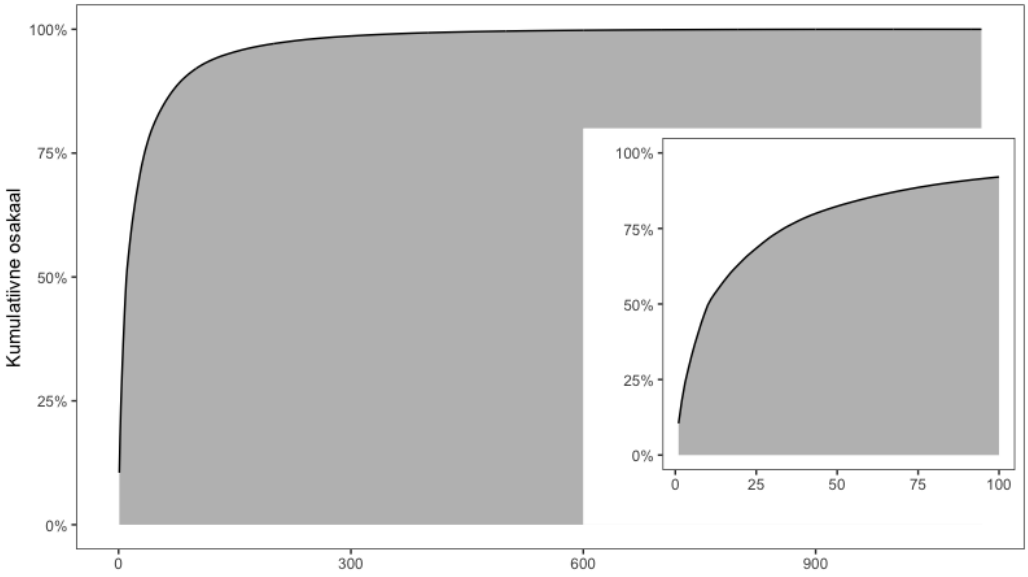
$$y_3 = 2,36 \cdot \text{liik } 1 - 2,97 \cdot \text{liik } 2 + 1 \cdot \text{liik } 3$$

Mittekorreleerituse tingimus on täidetud tänu omaväärtuste ortogonaalsuse omadusele. Esimene peakomponent kirjeldab $\frac{2,2}{2,2+0,62,0,16} = 0,74$ hajuvusest.

2 Ühe piimafarmi vasikate mikrobioomi mitmekesisuse analüüs

2.1 Andmed

Uuringu andmed koguti ühest suurest (u 800 lüpsilehmaga) Tartumaa piimafarmist. Uuringusse kaasati 66 lehmvasikat, kes sündisid ajavahemikus 17. oktoober 2018 kuni 31. märts 2019. Kõigilt vasikailt võeti rektaalse mikrobioomi proov 20 minuti jooksul peale sündi ning hiljem 1, 7, 10, 14, 21 ja 90 päeva vanuselt. Kaks vasikat surid esimesel elunädalal, mistõttu on neilt proovid vaid sünni- ja sellele järgnevalt elupäevalt. Erinevatel põhjustel jäid proovid võtmata osadelt vasikatelt vanemas eas, mistõttu on 10-päevastelt vasikatelt 62 ning 90-päevastelt vasikatelt 59 mikrobioomi proovi. Lisaks vasikatele võeti keskmiselt 7,3 päeva (miinimum üks ja maksimum 27 päeva, standardhälve 5,3 päeva) enne poegimist rektaalse mikrobioomi proovid ka 60 vasika emalt. Vasikaist 64 olid üksikud ja üks paar kaksikud lehmvasikad. Otse jämesoolest pärit proovidest eraldati DNA ning selle sekveneerimise tulemusena saadi info erinevate bakteriliikide esinemise kohta proovides neile liikidele omaste DNA-järjestuste lugemite arvude näol. Erinevatele proovidele vastavate DNA-järjestuste lugemite arvud sõltuvad paljudest asjaoludest, mistõttu ei ole liigi-spetsiifiliste lugemite arvud erinevates proovides otse võrreldavad. Seetõttu jagati liigi-spetsiifilised lugemite arvud proovides läbi proovide summaarsete lugemite arvudega, saades tulemuseks omavahel võrreldavad suhtarvud. Viimased moodustasidki käesoleva töö andmestiku. Kokku on andmestikus 500 mikrobioomiproovi andmed, neist 440 vasikatelt ja 60 lehmadelt. Iga proovi kohta on andmestikus 1114 erinevale bakteriliigile vastavate DNA-järjestuste lugemite arvud. 22 bakteriliigile vastavat DNA-d ei leidunud üheski proovis, seetõttu jäi lõplikuks liikide arvuks 1092. Liigid järjestati kogumahu järgi suuremast väiksemani ning koostati joonis kummulatiivse osakaalu kasvu vaatlemiseks (joonis 1). Joonise alumises paremas nurgas on toodud eraldi välja 100 dominantsema liigi kumulatiivne



Joonis 1: Mikrobioomi proovide liigipõhise koguse kumulaatiivne osakaal

osakaal kogu hulgast. Selgus, et proovides leidub väike hulk dominantseid liike, mis moodustavad enamiku kogu mikrobioomist. Seetõttu otsustati mikrobioomist eraldada tuumikmikrobioom, mis sisaldaks vaid laialdaselt levinud liike. Tuumikmikrobioomi eraldamisel lähtuti artiklis (Loch *et al.*, 2023) defineeritud põhimõttest, et tuumikmikrobioomi kuuluvad liigid, mis moodustavad vähemalt 0,1% proovist vähemalt 10% proovide puhul. Algsest 1092 liigist jäi tuumikmikrobioomi alles 144 liiki.

2.2 Statistilise analüüsi meetodid

Andmete analüüsimist alustatakse mikrobioomi α-mitmekesisuse mõõdikute arvutamisega. Mikrobioomi α-mitmekesisuse leidmise eesmärk on võrrelda, kuidas erinevad unikaalsete liikide arvud ja nende jaotused sõltuvalt looma vanusest. Liigirikkust, Shannoni entroopiat ja Simpsoni mitmekesisuse indeksit vaadeldakse nii tuumikmikrobioomis kui kogu mikrobioomis. Kolme erineva näidiku arvutamise eesmärk nii tuumikmikrobioomis kui kogu mikrobioomis on võrrelda kuivõrd koos-

kõlas on saadud tulemused. Kõigi α-mitmekesisuse näidikutega nii tuumikmikrobioomi kui ka kogu mikrobioomi andmetel viiakse läbi dispersioonanalüüs ja selle järel Tukey post-hoc testid tuvastamaks mikroobikoosluste mitmekesisuse poolest statistiliselt oluliselt erinevaid vanusegruppe.

Analüüsi jätkatakse mikrobioomi β-mitmekesisuse näidikute arvutamisega. β-mitmekesisuse leidmise eesmärk on vaadelda, kui sarnased on ühe vasika erinevatel vanustel antud proovid ning kui sarnased on need antud looma ema ja samuti farmi keskmise ema mikrobioomi prooviga. Farmi keskmise ema all mõeldakse farmi kõikide emade aritmeetilist keskmist proovi. Peamise mõõdikuna kasutatakse Bray-Curtise eripära, samas arvutatakse ka Eukleidilised kaugused, et võrrelda tulemuste kooskõla.

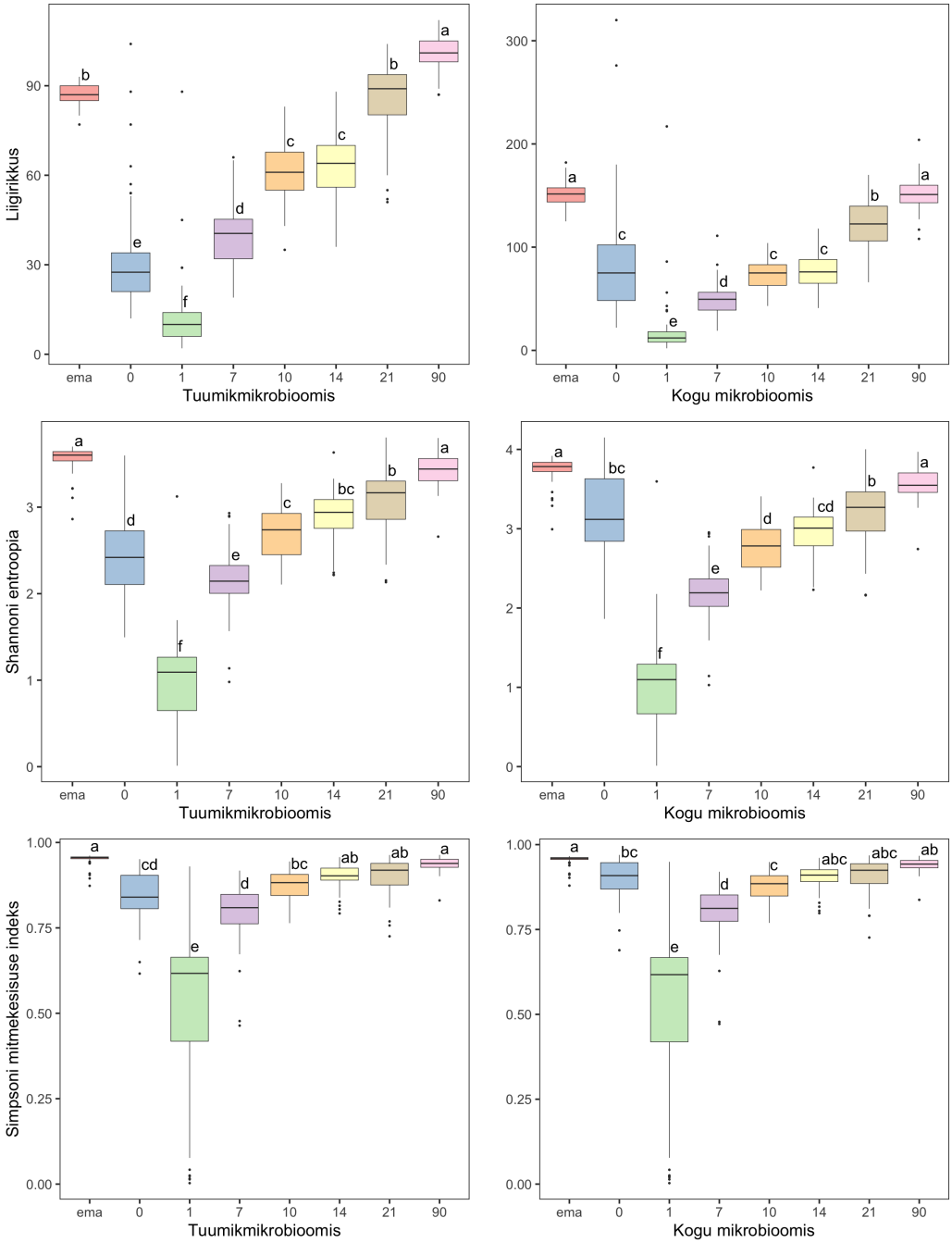
Mikrobioomi andmete analüüsi viimase osana viiakse läbi peakomponentanalüüs. Peakomponentanalüüsi eesmärk on leida proovides olevate liikide koguste lineaarkombinatsioone, mis kirjeldaksid võimalikult hästi kogu mikrobioomi.

Töö praktilise osa läbiviimiseks ning tulemuste graafiliseks kujutamiseks kasutatakse rakendustarkvara R (R Core Team, 2022). α- ja β-mitmekesisuse leidmisel kasutatakse paketti „vegan"(Oksanen *et al.*, 2022). Peakomponentanalüüsi läbi viimisel kasutatakse paketti „ade4"(Thioulouse *et al.*, 2018).

2.3 α-mitmekesisuse analüüs vasikate mikrobioomis

Uurimaks, kas soolestiku mikrobioomi mitmekesisus sõltub looma vanusest proovi võtmise hetkel, arvutati iga looma iga proovi kohta α-mitmekesisuse mõõdikud. Eraldi vaadeldi nii tuumikmikrobioomi kui ka kogu mikrobioomi α-mitmekesisust. Tulemused jagati gruppidesse vastavalt looma vanusele, kusjuures eraldi grupp moodustati ka emade mikrobioomi proovidest (Joonis 2).

Iga näidiku puhul teostati dispersioonanalüüs, et välja selgitada, kas gruppide vahel esineb statistiliselt olulisi erinevusi. Dispersioonanalüüs kinnitas, et iga mõõdiku korral leidub gruppidevahelisi erinevusi nii tuumikmikrobioomis kui ka terves mikrobioomis (kõik p-väärtused$<$0,01). Statistiliselt oluliste paarikaupa erinevuste



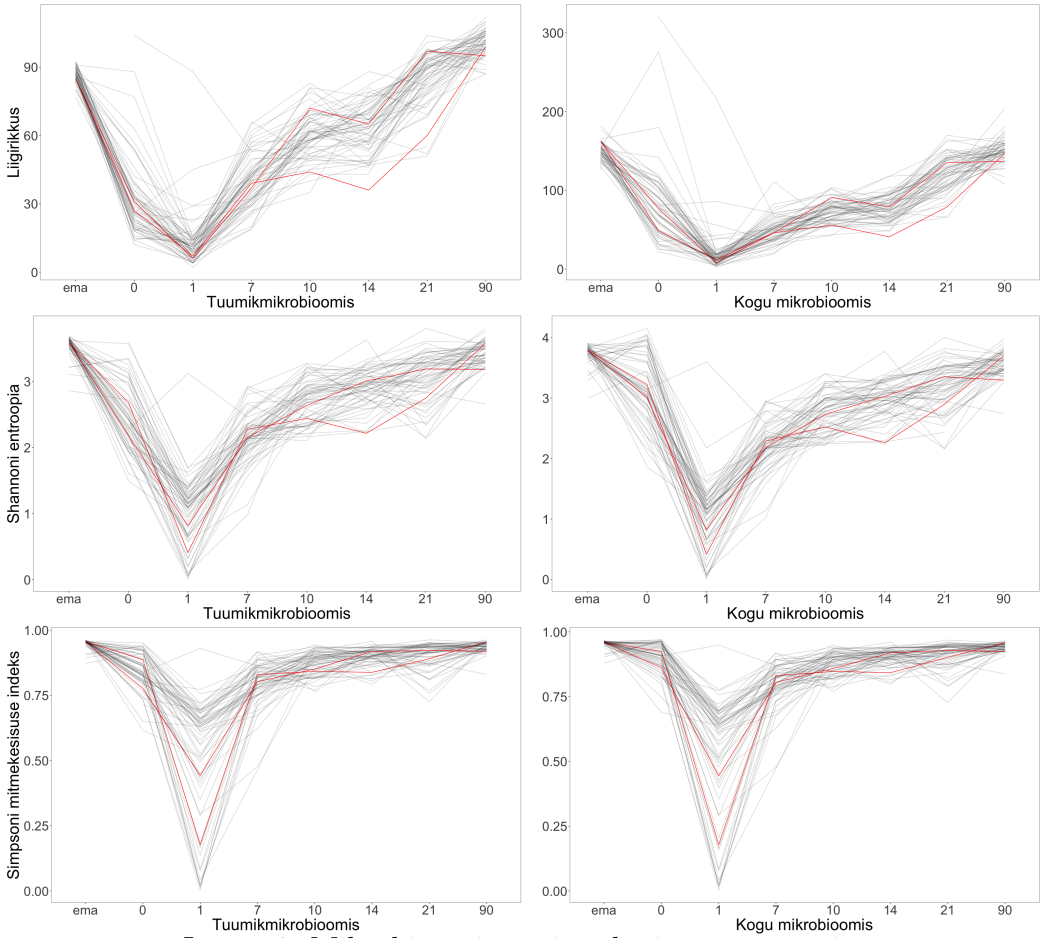
Joonis 2: α- mitmekesisus koos Tukey testiga

välja selgitamiseks viidi läbi Tukey test. Tulemused on nähtavad joonisel 2. Statistiliselt oluliselt erinevaks loetakse grupid, millel ei ole ühist tähte. Näiteks on erineva liigirikkusega elu esimesel ja seitsemendal päeval võetud tuumikmikrobioomi proovid. Samas ei tuvastatud erinevust 90 päeva vanuste vasikate ja emade tuumikmikrobioomi Shannoni entroopiate vahel. Shannoni entroopia ja Simpsoni mitmekesisuse indeks käituvad sarnaselt, kuid Simpsoni mitmekesisuse indeksi puhul on statistiliselt olulisi erinevusi veidi vähem. See viitab, et mikrobioomis domineerivad liigid on grupiti sarnasema jaotusega ning haruldasemate liikide jaotused erinevad rohkem. Selgub, et kui üldiselt vasika kasvades muutub looma mikrobioom aina mitmekesisemaks, siis sünnile järgneval päeval on vasika mikrobioom selgelt liigivaesem kui sündimise päeval. See viitab asjaolule, et osad lehma loote soolestikus elavad mikroobid ei ole välises keskkonnas elujõulised ja hukkuvad kiirelt. 90. päevaks on vasika soolestiku mikrobioom veidi liigirikkam, kuid üldiselt sarnase mitmekesisusega emade grupi mikrobioomiga. Siit ei saa siiski veel järeldada, et 90 päeva vanuste ja emade grupi mikrobioomid oleks sarnased. Teada on vaid, et liikide arv ja nende jaotus on sarnane. Seda, kas tegemist on valdavalt ka samade liikidega, hetkel teada ei ole. Selle uurimiseks tuleb vaadata β-mitmekesisust. Mõõdikute vahelised korrelatsioonid on arvutatud tabelis 6. Selgub, et korrelatsioonid on väga tugevad. Eriti silmatorkav on kogu mikrobioomi ja tuumikmikrobioomi samade mõõdikute vaheline korrelatsioon. Kõigil kolme mõõdiku puhul on korrelatsioonid üle 0,9, kusjuures Simpsoni mitmekesisuse indeksi puhul lausa 0,99. Siit on võimalik järeldada, et α-mitmekesisuse puhul ei ole olulist vahet, kas vaadelda tuumikmikrobioomi või kogu mikrobioomi.

Tabel 6: α-mitmekesisuse mõõdikute vaheline korrelatsioon

	Üldine liigrikkus	Üldine Shannon	Üldine Simpson	Tuumik liigirikkus	Tuumik Shannon	Tuumik Simpson
Üldine liigrikkus	1	0,84	0,65	0,91	0,85	0,66
Üldine Shannon	0,84	1	0,91	0,79	0,97	0,89
Üldine Simpson	0,65	0,91	1	0,66	0,89	0,99
Tuumik liigrikkus	0,91	0,79	0,66	1	0,88	0,7
Tuumik Shannon	0,85	0,97	0,89	0,88	1	0,91
Tuumik Simpson	0,66	0,89	0,99	0,7	0,91	1

Huvi pakub, kuidas toimub konkreetse vasika mikrobioomi mitmekesisuse areng ajas. Joonisel 3 on jälgitud iga looma soolestiku mikrobioomi arengut ajas. Joonise esimeses lõigus on iga vasikas kokku viidud oma emaga. Selgelt joonistub välja muster, et vasikas, kes oli nooremana mitmekesisema mikrobioomiga, on seda ka vanemana. Punase joonega on joonisel eraldi välja toodud andmestikus leidunud kaksikud. Liikide arvukus on kaksikutel olnud sarnane elu esimesel seitsmel päeval, seejärel on tekkinud erinevused, kuid 90. päevaks on kaksikute mikrobioomi liigiarvukused taas sarnased. See võib viidata asjaolule, et kaksikute mikrobioomide arengud on olnud erinevate kiirustega, kuid sarnase sihtpunktiga. Oluline on siiski märkida, et ühe kaksikupaari põhjal ei ole võimalik statistiliselt olulisi järeldusi teha.

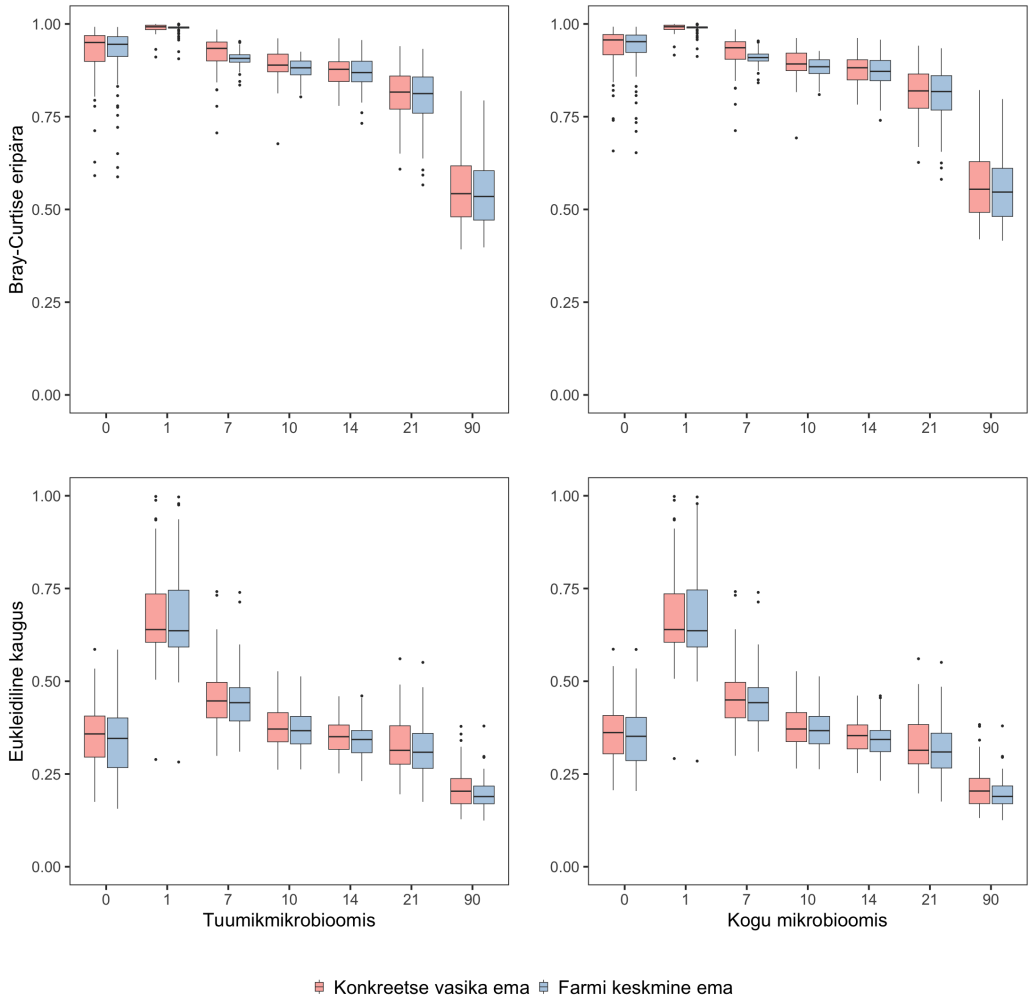


Joonis 3: Mikrobioomi α-mitmekesisuse areng ajas

2.4 β-mitmekesisuse analüüs vasikate mikrobioomis

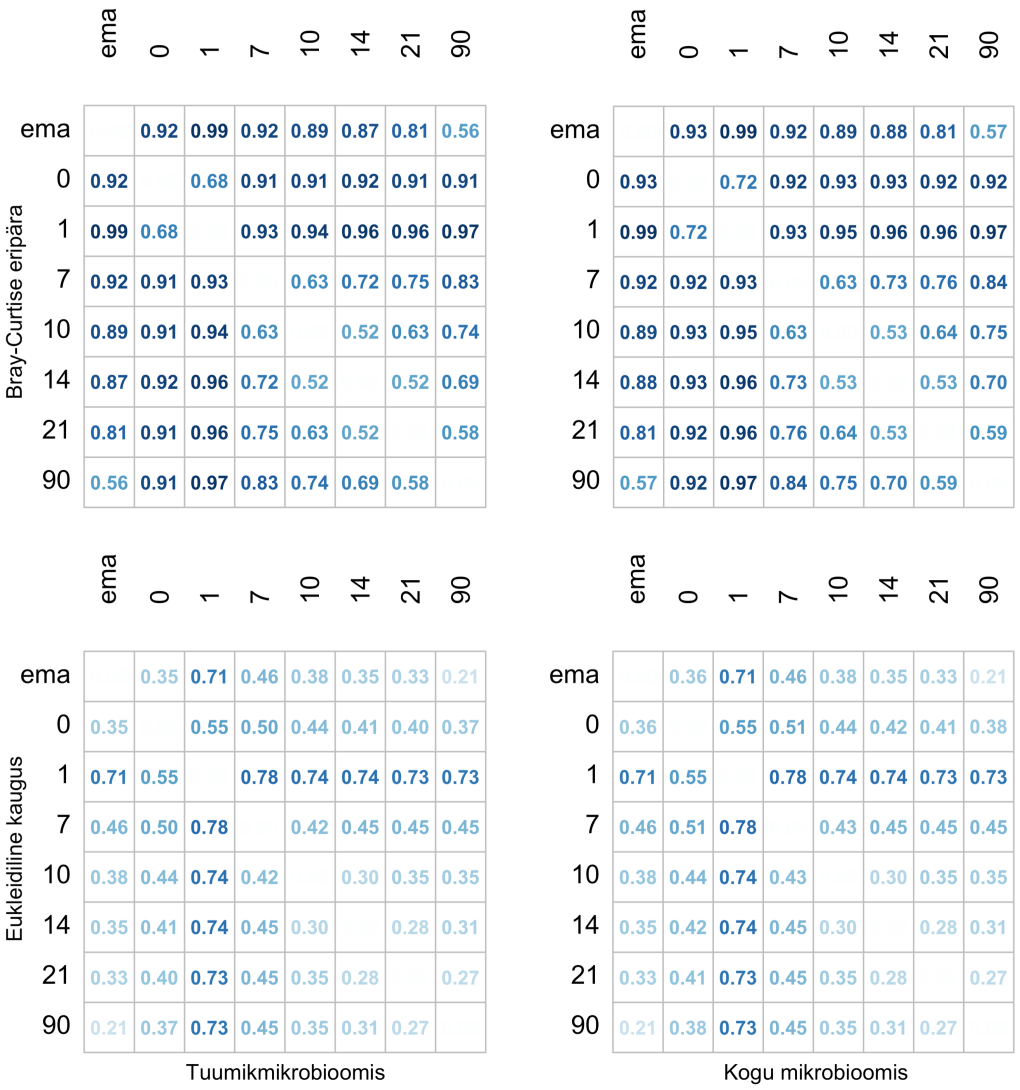
Vasikate ja nende emade mikrobioomide liigilise koosseisu võrdlemiseks kasutati β-mitmekesisust. Iga vasika igat proovi võrreldi tema ema prooviga ja kõikide farmi emade aritmeetilise keskmise prooviga. Saadud tulemused on nähtavad joonisel 4. Selgub, et meetodist ja vanusest olenemata ei ole vasikate mikrobioom sarnasem oma ema mikrobioomiga kui farmi keskmise ema mikrobioomiga. Bray-Curtise eripära põhjal saab väita, et ühegi proovi võtmise vanuse juures ei ole vasika ja ema mikrobioomid eriti sarnased. Seejuures ei ole vahet, kas vaadelda tuumikmikrobioomi või tervet mikrobioomi. Kõige sarnasemad on proovid võrreldes emaga 90 päeva vanuste vasikate puhul ning kõige erinevamad päev peale vasika sündi. Selgelt joo-

nistub välja muster, et pärast esimesest elupäeva hakkab vasika mikrobioom aina rohkem sarnanema ema omaga. Kahjuks lõppevad andmed 90 päeva juures ning saab vaid spekuleerida, kuidas toimunuks areng edasi. Eukleidline kaugus annab sarnaseid tulemusi Bray-Curtise eripäraga. Oluline erinevus ema ja järglaste mikrobioomide vahelisi erinevusi mõõtvates Bray-Curtise eripära ja Eukleidese kauguse väärtustes ilmneb nende varieeruvuses. Bray-Curtise eripära puhul on ühe päeva vanuste vasikate varieeruvus kõigist vanustest kõige ühesugusem. Eukleidilise kauguse puhul on selles vanuses loomade erinevus emast aga kõige varieeruvam.



Joonis 4: Vasika mikrobioomi liigilise koosseisu erinevus ema mikrobioomist

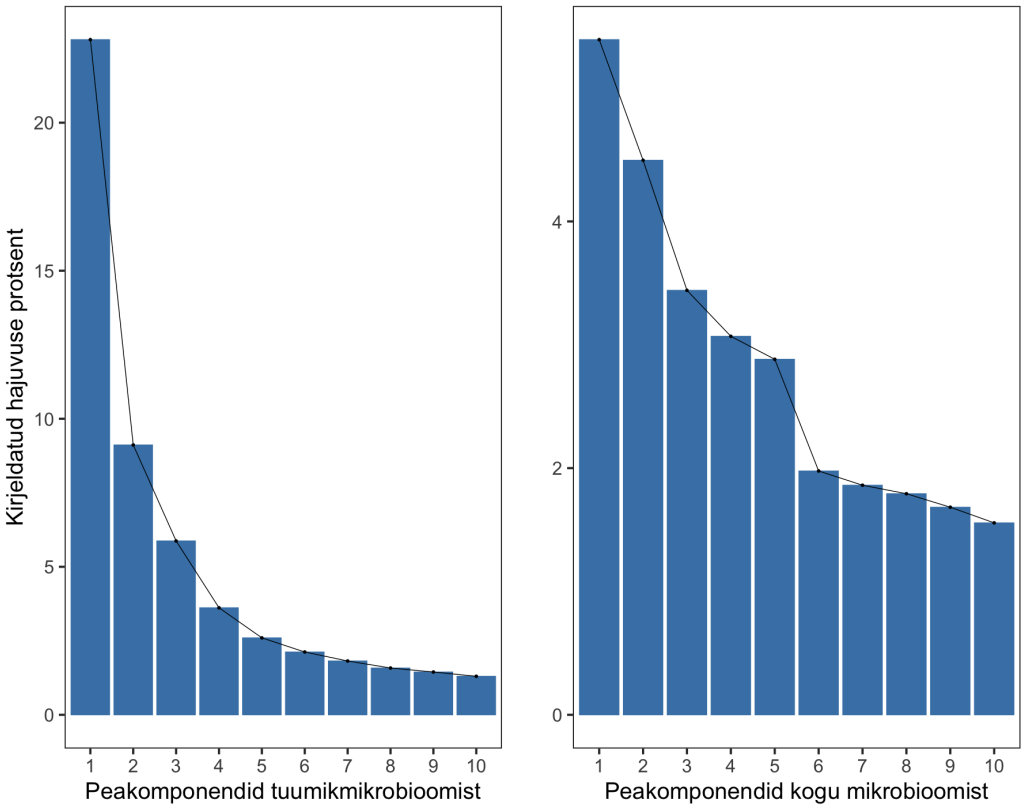
Huvi pakub, kui palju erinevad konkreetse vasika eri vanustel antud mikrobioomi proovid. Vasika eri vanustel antud mikrobioomi proovide võrdlemiseks, leiti iga proovi Bray-Curtise eripära ning Eukleidiline kaugus võrreldes sama vasika teiste proovidega. Taas vaadati eraldi tuumikmikrobioomi ja kogu mikrobioomi. Niimoodi saadi iga looma kohta neli 8×8 tabelit. Seejärel leiti iga tabeli aritmeetiline keskmine üle kõigi loomade. Saadud tulemused on nähtavad joonisel 5. Ootuspäraselt on üksteisele ajaliselt lähemal olevad proovid keskmiselt sarnasemad. Esimesel elupäeval antud proovid erinevad selgelt kõige rohkem teistest proovidest. Sellest saab järeldada, et esimesel elunädalal oleks pidanud võtma proove tihedamini, sest mikrobioomi muutused sel perioodil on väga kiired.



Joonis 5: Vasika mikrobioomi keskmine erinevus eri vanustes

2.5 Vasikate mikrobioomiandmete peakomponentanalüüs

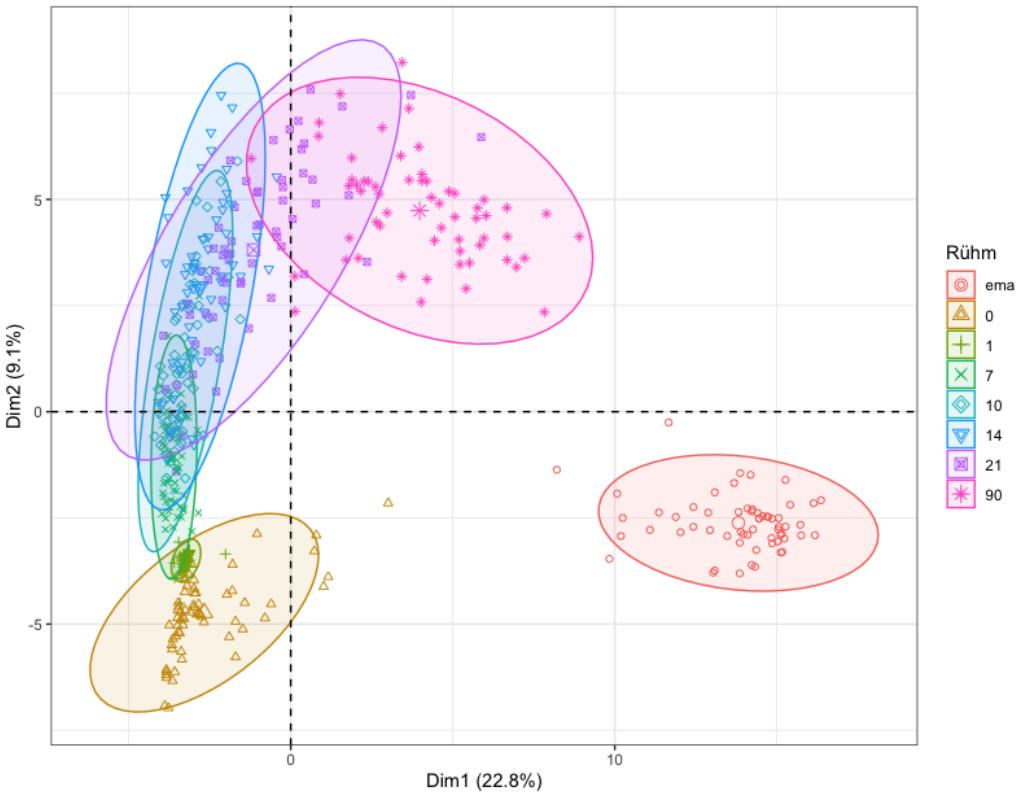
Enne peakomponentanalüüsi läbiviimist normeeriti proovide väärtused nii, et keskväärtus oleks 1 ja standardhälve 0. Algselt teostati analüüs nii tuumikmikrobioomi kui ka kogu mikrobioomi puhul. Esimese 10 peakomponendi poolt kirjeldatud hajuvused on toodud joonisel 6. Peakomponentanalüüsi teostamisel tuli selgelt välja tuumikmikrobioomi eraldamise vajadus. Tuumikmikrobioomi puhul kirjeldavad



Joonis 6: Tähtsamate peakomponentide poolt kirjeldatud hajuvus tuumikmikrobioomis ja kogu mikrobioomis

esimesed kaks peakomponenti koos 32% hajuvusest, samas kui kogu mikrobioomis suudavad esimesed kaks peakomponenti kirjeldada vaid 10% hajuvusest. Seetõttu otsustati peakomponentanalüüsi jätkata vaid tuumikmikrobioomiga.

Joonisel 7 on arvutatud iga proovi esimese kahe peakomponendi väärtused, proovid on grupeeritud vasika vanuse järgi. Eraldi grupi moodustavad emade proovid. Joonise interpreteerimisel tuleb arvestada, et x-telg kirjeldab $\frac{22,8\%}{9,1\%} = 2,5$ korda rohkem y-teljest. See tähendab, et ühe ühiku pikkune vahemaa x-teljel on ekvivalentne 2,5 ühiku pikkusele vahemaale y-teljel. Jooniselt joonistuvad selgelt välja rühmad. Seega on samas vanuses loomade mikrobioomi proovid märkimisväärselt sarnasemad kui konkreetse looma eri vanustes antud mikrobioomi proovid. Osutub selgelt, et juba 20 minuti vanuste vasikate mikrobioom on oluliselt erinev nen-



Joonis 7: Proovide esimese kahe peakomponedi väärtused

de ema mikrobioomist. Ühe päeva vanuste vasikate esimese kahe peakomponendi väärtused on väga väikse hajuvusega. Sellest saab järeldada, et antud vanuses on vasikate mikrobioomid väga sarnased ning seega on mõju emalt väike või puudub üldse. Jooniselt on selgelt näha, et vasika kasvades tema mikrobioom hakkab aina enam sarnanema emade rühma mikrobioomiga. Peakomponentanalüüs kinnitab joonise 4 põhjal tehtud järeldust, et 90. päevaks ei ole mikrobioom veel lõplikult välja kujunenud.

Kokkuvõte

Käesoleva bakalaureusetöö eesmärgiks oli uurida erinevaid võimalusi, kuidas analüüsida mikrobioomi andmeid. Töö esimeses osas anti ülevaade mikrobioomi andmete analüüsimise statistilisest metodoloogiast. Töö teises pooles analüüsiti ühe Tartumaa suure piimafarmi vasikate mikrobioomi andmeid. Analüüs hõlmas α- ja β-mitmekesisuse mõõdikute leidmist ning peakomponentanalüüsi läbi viimist.

α-mitmekesisuse analüüsimisel selgus, et vasika mikrobioom on kõige liigivaesem looma sünnipäevale järgneval elupäeval. β-mitmekesisuse analüüsimisel ilmnes, et vasika mikrobioom ei ole sarnasem enda emaga kui sama farmi keskmise emaga. Peakomponentanalüüsist tulenes, et 90. päevaks ei ole veise mikrobioom veel välja kujunenud. Seetõttu tuleks tulevastes uuringutes võimalusel kaasata ka vanemasse ikka jõudnud veiseid.

α- ja β-mitmekesisuse puhul ei erinenud tulemused oluliselt sellest, kas vaadeldi vaid tuumikmikrobioomi või kogu mikrobioomi. Peakomponentanalüüsi korral katsid olulisemad peakomponendid tuumikmikrobioomi puhul märkimisväärtselt suurema osa hajuvusest kui kogu mikrobioomi vaatlemise korral.

Kasutatud allikad

Johnson, R.A. ja D. W. Wichern (2007). *Applied Multivariate Statistical Analysis, 6th edn.* Pearson Prentice Hall.

Kers, J. G. ja E. Saccenti (2021). "The Power of Microbiome Studies: Some Considerations on Which Alpha and Beta Metrics to Use and How to Report Results". *Frontiers in Microbiology* 796025.12.

Loch, M., E. Dorbek-Kolin, A. Husso, T. Pessa-Morikawa, T. Niine, T. Kaart, K. Mõtus, M. Niku ja T. Orro (2023). "Associations of neonatal faecal microbiota with inflammatory markers, growth rates and first lactation performance of dairy cows". Publitseerimiseks saadetud käsikiri.

Oksanen, J. (2022). *Vegan: ecological diversity.* URL: https://cran.r-project.org/web/packages/vegan/vignettes/diversity-vegan.pdf (vaadatud 09.05.2023).

Oksanen, J., G.L. Simpson, F.G. Blanchet, R. Kindt, P. Legendre, P.R. Minchin, R.B. O'Hara, P. Solymos, M.H.H. Stevens, E. Szoecs, H. Wagner, M. Barbour, M. Bedward, B Bolker, D. Borcard, G. Carvalho, M. Chirico, M. De Cáceres, S. Durand, H.B.A Evangelista, R. FitzJohn, M. Friendly, B. Furneaux, G. Hannigan, M.O. Hill, L. Lahti, D. McGlinn, M.H. Ouellette, E. Ribeiro Cunha, T. Smith, A. Stier, C.J.F. Ter Braak ja J. Weedon (2022). *vegan: Community Ecology Package.* R package version 2.6-4. URL: https://CRAN.R-project.org/package=vegan (vaadatud 10.04.2023).

Orloci, L. (1978). *Multivariate analysis in vegetation research.* URL: https://scholar.google.com/scholar_lookup?title=Multivariate%20Analysis%20in%20Vegetation%20Research&publication_year=1978&author=L.%20OrlÃşci (vaadatud 09.05.2023).

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/ (vaadatud 09.05.2023).

Slanzon, G.S., B.J. Ridenhour, D.A. William Moore, M. Sischo, L.M. Parrish, S.C. Trombetta ja C.S. McConnel (2022). "Fecal microbiome profiles of neonatal dairy calves with varying severities of gastrointestinal disease". *PloS one* 17.

Thioulouse, J., S. Dray, A.B. Dufour, A. Siberchicot, T. Jombart ja S. Pavoine (2018). *Multivariate Analysis of Ecological Data with ade4*. Springer. DOI: 10.1007/978-1-4939-8850-1.

Tiit, E.-M. ja M. Viil (1992). *Andmeanalüüs personaalarvutil programmipaki Statgraphics abil*. Tartu Ülikool.

Zhu, H., M. Yang, J.J. Loor, A. Elolimy, L. Li, C. Xu, W. Wang, S. Yin ja Y. Qu (2021). "Analysis of Cow-Calf Microbiome Transfer Routes and Microbiome Diversity in the Newborn Holstein Dairy Calf Hindgut". *Frontiers in Nutrition* 8.

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Mikk Tomson,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose „Vasikate seedetrakti mikrobioomi mitmekesisus ja areng", mille juhendaja on Tanel Kaart, 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 Commonsi litsentsiga CC BY NC ND 4.0, 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 intellektuaaloman-di ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Mikk Tomson
09.05.2023