

**Mikrobiom elemzés során alkalmazott 16S rRNS gén V1-V3 és V4 amplikon
szekvenálási módszerek összehasonlítása**

36. OTDK Biológia Szekció

Henkrich Miksa

Biológia BSc IV. évfolyam

A dolgozat írásakor Biológia BSc III. évfolyam

Járay Tamás

Doktori iskola I. évfolyam

A dolgozat írásakor Biológia MSc II. évfolyam

Témavezetők:

dr. Tombácz Dóra egyetemi docens

Gulyás Gábor PhD-hallgató

SZTE SZAOK Orvosi Biológiai Intézet

GeMiNI munkacsoport

Szeged

2022

Rövidítésjegyzék

16S: 16 Svedberg egység

ASV: amplitudon szekvencia variáns

BARF: Csont és Nyers Hús Étrend /Biológiailag Helyénvaló Nyers Étrend (Bones and Raw Food Diet/Biologically Appropriate Raw Food Diet)

DMD: Duna-menti Dumás Kennel

DNS/DNA: dezoxiribonukleinsav

dNTP: dezoxiribonukleotid-trifoszfát

HTML: HyperText Markup Language

mRNS: hírvivő (messenger) RNS

ONT: Oxford Nanopore Technologies

OTU: operációs taxonómiai egység

PCR: polimeráz láncreakció

RNS: ribonukleinsav

rRNS: riboszomális RNS

V1-V9: a 16S riboszomális RNS 9 hipervariábilis régiójának rövidítései

Tartalomjegyzék

Rövidítésjegyzék	1
Tartalomjegyzék	2
Bevezető	3
1. A gazda-mikrobiom-szuperorganizmus	3
2. Az elhízás és a mikrobiom kapcsolatának kutatása	4
3. A 16S rRNS gén V régió amplikon szekvenálás kihívásai	5
4. A metagenomika bevett vizsgálati megközelítései és munkafolyamatai	6
4.1 Az újgenerációs szekvenálás	7
4.2 Rövid leolvasású szekvenálás Illumina MiSeq készülékkel	8
4.3 Célzott amplikon szekvenálás és a riboszomális V régiók	10
4.4 A 16S rRNS amplikon szekvenálási adatok kiértékelése bioinformatikai módszerekkel	11
Célkitűzés	14
Anyagok és módszerek	15
1. Modellállatok és mintagyűjtés	15
1.1 Vizsgálati alanyok és étrendjük	15
1.2 Mintavételezési időpontok, minták kezelése	16
2. DNS izolálás, könyvtárkészítés és szekvenálás	17
2.1 DNS-izolálás kutya székletmintából	17
2.2 Amplikon szekvenálási könyvtárkészítés	18
2.3 16S rRNS V1-V3 és V4 amplikon könyvtárak szekvenálása	18
3. A szekvenálási adatok bioinformatikai és manuális elemzése	18
3.1 Amplikon szekvencia variánsok taxonómiai azonosítása, az adatok vizualizációja	18
3.2 Manuális adatelemzés	19
4. A munkafolyamat lépései összefoglalva	20
Eredmények és értékelésük	22
1. Azonosított ASV-k	22
2. Diverzitásindexek	23
3. Klaszter dendrogramok	26
4. Eredményeink összefoglalása	30
Összefoglalás	31
Irodalomjegyzék	32
Ábrák forrásai	34
Köszönetnyilvánítás	36

Bevezető

1. A gazda-mikrobiom-szuperorganizmus

Az utóbbi években komoly előrelépések történtek a mikrobiomika, azaz a többsejtű gazdaszervezetekkel szimbiózisban élő mikroorganizmusok kutatásának területén. Az újgenerációs szekvenálási eljárásokon alapuló módszerekkel a vizsgált baktériumközösségek nagy pontosságú kompozíciós és funkcionális vizsgálata vihető végbe. A metagenom, azaz egy vizsgált mikrobiális közösség teljes génállománya marker gének vagy akár a génexpresszió szintjén vizsgálható a korszerű technológiáknak köszönhetően. A mikrobiom-kutatás talán legjelentősebb eredményei közé tartoznak azon felfedezések, melyek az intesztinális mikrobiom vagy bélflóra humán egészségügyi relevanciáját tárják fel.

Sender és munkatársai (2016) becslései nyomán egy átlagos, 70 kilogrammos felnőtt ember körülbelül $3,8 \cdot 10^{13}$ bakteriális sejtnek ad otthont, azaz a humán és a vele együtt élő prokarióta sejtek száma hasonló. A legtöbb baktérium a vastagbélben található. Itt a béltartalom milliliterenként akár 10^{11} sejtet tartalmazhat, a vékonybél baktériumtartalma $10^3 - 10^8$ sejt/ml között változik disztálisan nagyobb bakteriális abundanciával, a kolonizáló mikrobákban relatíve szegény gyomor pedig hozzávetőleg $10^3 - 10^4$ mikroorganizmusnak ad otthont (Sender és mtsai. 2016). A humán bél mikrobiom megközelítőleg 500-1000 fajt számlál, melyek körülbelül kilenctized része az Actinobacteria, Bacteroides, Firmicutes és Proteobacteria törzsek valamelyikébe tartozik (Thursby, Juge 2017). Az emberrel közösséget alkotó mikrobák kollektív genomja, metagenomja a humán genom százszorosát is kiteheti, hiszen egy-egy baktérium gének ezreit hordozza, és a prokarióták között a fajon belüli genetikai változatosság (bakteriális *strain*, faj alatti törzs szinten) igen nagy lehet. Li és Jia 2014-ben publikált, közel 1300 humán minta vizsgálata alapján összeállított génkatalógusa a humán metagenom majdnem tízmillió génjét azonosította.

A Human Microbiome Project szerzői nyomán a mikrobiom mag és perifériás (*core* és *accessory*) egységeit különböztetjük meg (Turnbaugh és mtsai. 2013). A mikrobiomra nagyfokú funkcionális redundancia jellemző, azaz egy adott fiziológiás - metabolikus funkció betöltéséhez szükséges génkészlettel egy egészséges mikrobiom számos tagja rendelkezik. Így egy fontos funkciót betöltő faj valamilyen hatásra történő eltűnése nem feltétlenül jár az adott funkció kiesésével, hiszen azt egy hasonló metabolikus útvonalakkal, életműködésekkel bíró faj fogja ellátni a továbbiakban. Az egy adott mikrobiomra jellemző fő funkciókat ellátni képes génkészletet, illetve a génkészletet hordozó fajok összességét nevezzük mag mikrobiomnak. A mag mikrobiom az idő és más

környezeti tényezők változásainak fényében nagyobb állandóságot mutat, mint a kevésbé esszenciális, összetételét tekintve nagyobb mértékben fluktuáló perifériás mikrobiom.

A gasztrointesztinális traktus rezidens mikrobái komoly szerepet játszanak a gazda fiziológiás és patológiás működéseiben a bevitt tápanyagok módosítása, különféle bioaktív anyagok szintézise és az immunrendszer folyamatos modulálása folytán. Egy egészséges és reziliens mikrobiom kulcsa lehet a gazdaszervezet jó állapotának, ugyanakkor egyre több krónikus betegség hátterében sejlik fel az intesztinális diszbiózis mint potenciális kiváltó ok, vagy közvetítő tényező a kedvezőtlen környezeti hatások és a nemkívánt fiziológiás elváltozások, betegségek között (DeGruttola és mtsai. 2016, Wilkins, Monga, Miller 2019, Vilela de Oliviera 2019). A diszbiotikus, azaz a bélflóra egyensúlyának felborulásával járó állapotokra jellemző a bakteriális közösség diverzitásának és funkcionális redundanciájának csökkenése, a bélhám fokozott átteresztőképessége, baktériumok és egyéb, a bél lumenében található anyagok transzlokációja a bélből a szervezet belső környezetébe, az immunrendszer működésének megváltozása és gyulladásos folyamatok aktiválódása (Vilela de Oliviera 2019).

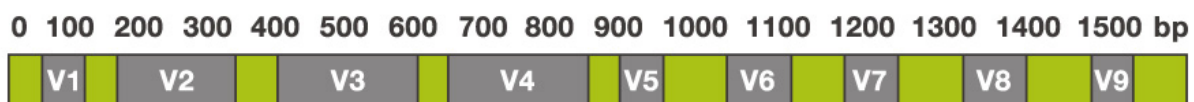
2. Az elhízás és a mikrobiom kapcsolatának kutatása

Az elhízásban és az ahhoz kapcsolódó számos társbetegségben az alacsony szintű krónikus gyulladás és a diszbiotikus mikrobiom szerepe erősen indikált (Iacobini és mtsai. 2018, Liébana-García és mtsai. 2018). A terület eredményeinek részletezése módszertani fókuszú dolgozatomban témáján túlmutat, így a következő sorokban kutatócsoportunk ide vonatkozó célkitűzéseit mutatom be, melyek a mikrobiom elhízásban, inzulinrezisztenciában és kettes típusú diabéteszben betöltött szerepének, illetve a mikrobiom étrendi szénhidráttartalomtól való függésének megértése. Munkánk során elhízásra, inzulinrezisztenciára és kettes típusú diabéteszre eltérő egészségi mutatókkal bíró, eltérő szénhidrátbevitelt előíró diétás-és életmódprogramokat követő humán páciensek, illetve fajtatiszta, ismert rokonsági viszonyú, a humán mikrobiom kiváló modellorganizmusául szolgáló (Coelho és mtsai. 2018) pumikutyák székletmintáinak longitudinális elemzését végezzük különböző második és harmadik generációs szekvenálási eljárásokkal. Az alkalmazott szekvenálási eljárásokkal a mikrobiom metagenom és metatranszkriptom mellett a viromot, illetve a gazdaszervezet építélejtjeinek metilomját és RNS-profilját is vizsgáljuk - mindazonáltal kutatómunkánk egyik kulcsfontosságú eljárása a minta baktériumfajainak azonosítására, így asszociatív kapcsolatok megállapítására alkalmas 16S rRNS gén ampliconjainak szekvenálása.

A prokarióta 16S riboszomális RNS-t kódoló gén (*1. ábra*) speciális felépítéséből kifolyólag nagyban alkalmas a gént hordozó mikroba taxonómiai azonosítására. A génben hiperkonzervatív, távoli rokon fajok között is hasonló szekvenciájú régiók között hipervariábilis, közeli rokon fajok között is eltérő bázissorrendű régiók találhatóak. A gén összesen 9 hipervariábilis régiót hordoz, melyek jelölése V1-V9. A variábilis régiók az őket közrefogó konzervatív régiókra tervezett univerzális PCR-primerekkel amplifikálhatók, a létrehozott amplikonok pedig, méretüktől függően rövid vagy hosszú leolvasású módszerrel, megszekvenálhatók, a szekvenciák eredete pedig bioinformatikai elemzés útján, referencia adatbázisok segítségével megállapítható. Magyarán, egy baktériumokat ismeretlen összetételben tartalmazó minta kompozíciós elemzése elvégezhető a 16S rRNS gén tetszőleges V régióinak amplikon szekvenálásával, és a módszer a fellelt fajok abundanciaértékét is jelzi. Ezáltal egy adott mintaadó alany mikrobiom-összetételéről nyerhetünk betekintést, és összefüggéseket ismerhetünk fel a mikrobiom-összetétel és az étrend, az egészségi állapot vagy akár a gazda genetikája között.

3. A 16S rRNS gén V régió amplikon szekvenálás kihívásai

Mivel összesen 9 vizsgálható, egymástól eltérő V régiót ismerünk, a régió választása befolyásolhatja a kapott eredményeket: eltérő V régió amplikon szekvenciák részben eltérő összetételi adatokat mutathatnak a vizsgált közösség összetételéről és diverzitásáról.



1. ábra: A 16S rRNS gén régiói. Konzervatív régiók, variábilis régiók

A különböző V régiók szekvenciadiverzitásának mértéke egymástól eltér, illetve a különböző taxonok különböző mértékben halmoznak fel mutációkat a gén egyes V régióiban (Chakravorty és mtsai. 2007). Több tanulmány foglalkozott a különböző V régió markerek taxonómiai azonosításra való alkalmasságának értékelésével (Chakravorty és mtsai. 2007, Bukin és mtsai. 2019, Zhang és mtsai. 2018, Kerrigan és mtsai. 2019). Ezen tanulmányok eredményei megerősítik a feltételezést, hogy a választott V régió befolyásolja a kapott eredményeket. A különböző V régiók különböző mértékben alkalmasak az egyes taxonok azonosítására, az egyik, 2007-es tanulmány (Chakravorty és mtsai. 2007) szerint az is régiófüggő, hogy milyen mélységig, azaz nemzetség vagy akár faj szintig is lehetséges-e az azonosítás.

A kilenc hipervariábilis régió a 16S rRNS háromdimenziós szerkezetében eltérő helyet foglal el és eltérő szerepet tölt be, melynek következtében mutációs hajlandóságuk is eltér (Bukin és mtsai. 2019, Schuwirth és mtsai. 2005, Schlutzen és mtsai. 2000). A V4, V5 és V6 a tRNS-ekkel és a 23S rRNS-sel való interakcióban vesznek részt, ennek következtében nagyobb strukturális stabilitás jellemző rájuk. A legvariábilisabb V régiók a 16S rRNS strukturális integritásának biztosításában részt vevő V2 és V8; a jelenleg kevésbé ismert szerepekkel bíró V3 és V7 régiók nukleotidjainak mutációs gyakorisága a többi régióhoz viszonyítva közepes. Minél konzervatívabb egy V régió, annál inkább igaz rá, hogy a közeli rokon fajok között hasonló a szekvenciája, és alapvetően a magasabb fokú taxonómiai azonosításra alkalmas. Ezen ismeretek alapján feltételezhető, hogy a V4 amplikon inkább törzs és osztály, a V2 és V3 régiókat tartalmazó V1-V3 amplikon ennél alacsonyabb szinteken történő taxonómiai azonosításra ideális.

Látható, hogy egy kutatás során alkalmazott V régió megfelelő körültekintéssel történő kiválasztása kulcsfontosságú a kutatás eredményeinek szempontjából, azonban az általunk áttekintett referencia szakirodalom alapján nincs egyértelmű tudományos konszenzus a bakteriális közösségek kompozíciós vizsgálatára legalkalmasabbnak tartott V régióval kapcsolatban.

Kutatócsoportunk a V1-V3 régió (pozíció: 69-497 nukleotid) és a V4 régió (pozíció: 576-682 nukleotid) amplikonokat vizsgálta, ezen amplikonok szekvenciái alapján igyekeztünk megállapítani, hogy melyik régió milyen képet ad a vizsgált mikrobiom-mintákról, illetve, hogy a két módszer közül potenciálisan melyik fedi le jobban a vizsgált mikrobiomok valós összetételét.

4. A metagenomika bevett vizsgálati megközelítései és munkafolyamatai

A fent részletezett precíziós szekvenálási eljárások kivitelezhetősége és a mikrobiom-kutatásfellendülése a molekuláris biológiai, elsősorban szekvenálási technikák 21. században tapasztalt drasztikus fejlődésének köszönhető. Az újgenerációs szekvenálási módszerek megjelenése és gazdaságossá válása előtt a bélflóra összetételét vizsgálni nehézkes volt: a korábbi évtizedek kutatói szinte kizárólag kulturomikai technikákra támaszkodhattak. Azonban a bélflóra tagjai, különösen a vastagbél-baktériumok, nagy százalékban anaerob vagy mikroaerofil anyagcserét folytatnak, így sztenderd laboratóriumi körülmények között nagy részük nem vagy csak körülményesen tenyészthető.

Kortárs módszerekkel viszont nagy tisztaságú DNS és RNS izolálható széklet- vagy kimuszmintából, a nukleinsavak szekvenciája pedig könnyen és viszonylag olcsón meghatározható újgenerációs vagy masszív paralell szekvenálási eljárásokkal. Ez széles távlatokat nyit meg a

kutatás számára: nagy fajok közötti variabilitást prezentáló genomi régiók vizsgálatával nagy pontossággal azonosíthatók a vizsgált mintában található taxonok, genomikai markerek alapján megkülönböztethetők a metagenom prokarióta, eukarióta, vírus és gazda eredetű szekvenciái, átfedő szekvenciák egymásra illesztésével kiterjedt genomi régiók és teljes genomok állíthatók össze, RNS-szekvenciák vizsgálatával transzkriptomikai elemzés végezhető, mely metabolomikai és funkcionális következtetéseknek ad teret, metilom-szekvenálással pedig még bizonyos epigenetikai mintázatok is azonosíthatók. Ilyen módszertani arzenállal felfegyverkezve a kortárs mikrobiomikai kutatás nem csupán kompozíciós, de igen komplex funkcionális elemzések elvégzésére is képes.

A továbbiakban bemutatom a metagenom-szekvenálás jelen projektben alkalmazott módszereinek elméleti hátterét.

4.1 Az újgenerációs szekvenálás

Az újgenerációs szekvenálási eljárásokat masszív parallel szekvenálásnak is nevezik, mert a korábban elterjedt Sanger-szekvenálással szemben igen nagyszámú minta párhuzamos leolvasását teszik lehetővé. Ezen platformok nagy része, így a csoportunk által alkalmazott Illumina MiSeq (Ravi, Walton, Khosroheidari 2018) szekvenátor a Sanger-szekvenáláshoz hasonlóan szintézisalapon szekvenál, melynek során a szekvenált nukleinsav egyik szála templátként szolgál a komplementer szál in vitro szintéziséhez. A szintézis fluoreszcensen jelölt nukleotidokkal történik, az egyes nukleotidok beépülési helyén egyedi színjelenség detektálható, a detektált színjelenségek alapján pedig megállapítható a szekvenált molekula bázissorrendje. Az újgenerációs szekvenátorok esetében a fluoreszcens felvillanások felfogása minden egyes szintetikus lépés után megtörténik, így a Sanger-szekvenálással szemben egyszerre igen nagyszámú DNS-molekula szekvenálható meg az újgenerációs platformokon. Ezen készülékek emellett gyorsabban is dolgoznak a Sanger-szekvenátoroknál, mivel a potenciálisan beépülő 4 nukleotid egyszerre adható a mintához, a munkafolyamatot az egyes nukleotidok felvitele közt mosási lépések nem lassítják. A szintézisalapú szekvenátorok által generált leolvasások hossza azonban csupán néhány száz bázispár. Ennél hosszabb, akár tízezer bázispár nagyságrendű readok hosszú leolvasású szekvenálással generálhatók, a módszer azonban nem képezi jelen pályamunka tárgyát.

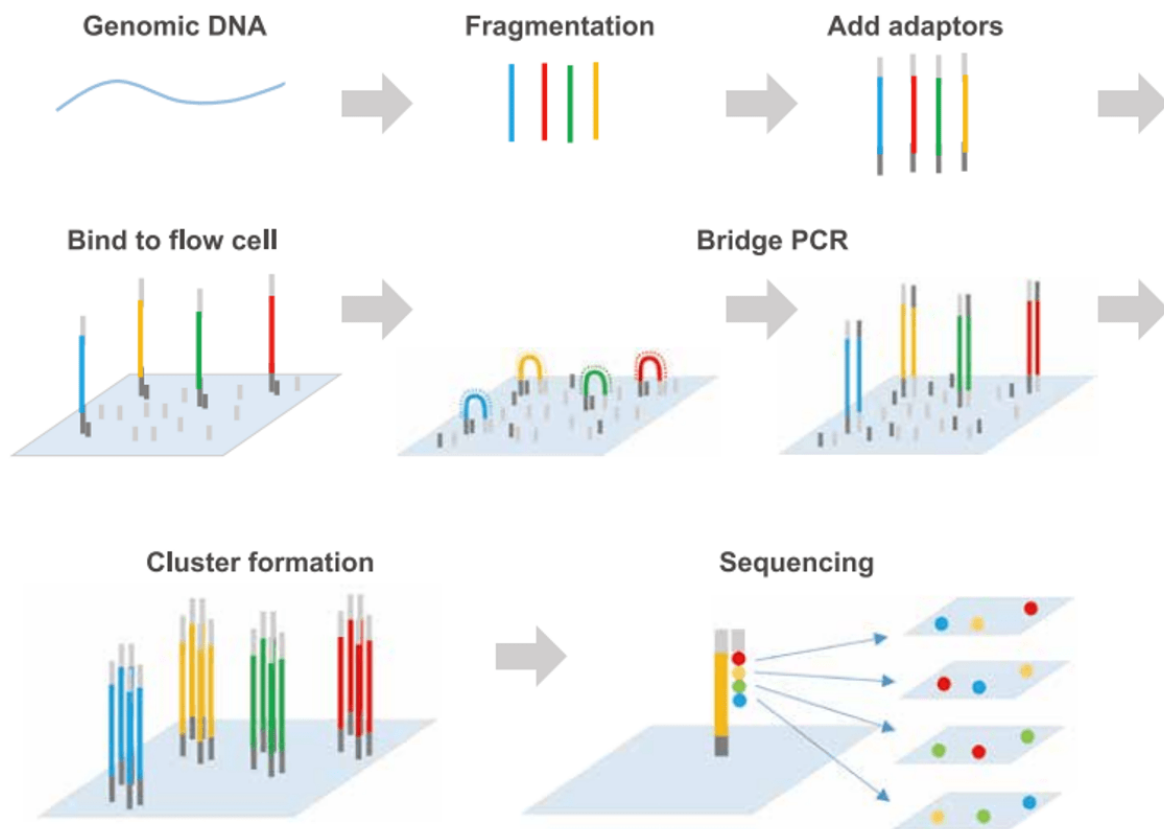
4.2 Rövid leolvasású szekvenálás Illumina MiSeq készülékkel

Az Illumina MiSeq rövid leolvasású, szintézisalapú masszív parallel szekvenátor. Az eszköz alapegysége az akrilamiddal burkolt üveg flow cell, a szekvenálásra előkészített DNS-mintát ide

kell felvinni. A flow cell felületén nanométeres mélyedésekben szintetikus oligonukleotid próbák rögzülnek, melyek a szekvenálási könyvtár elemeit a flow cellhez képesek horgonyozni.

A szekvenálási könyvtár, azaz a leolvasásra megfelelően előkészített DNS-molekulák összességének elemei kettős szálúak és egységesen néhány 10, legfeljebb néhány 100 bázispár hosszúak. Végeikhez úgynevezett adapter van ligálva: ez egy három funkcionális régióból álló szintetikus oligonukleotid, mely tartalmaz egy, a flow cellre rögzített próbák valamelyikével komplementer szekvenciát, egy egyedi azonosítót vagy barcode-ot, valamint a szekvenálást végző polimeráz enzim primerjének bekötését segítő szekvencia is az adapter részét képezi.

A könyvtár elemei a flow cellre kerülve egy-egy oligonukleotid próbához horgonyozódnak, ezt követően lezajlik egy in situ polimeráz láncreakció, melyet klasztergenerációnak (2. ábra) nevezünk. A klasztergeneráció során minden egyes felvitt DNS-molekuláról körülbelül ezer másolat képződik, így minden eltérő DNS-molekula megszekvenálása körülbelül ezerszeres mennyiségben történik. A különböző klaszterek térben elkülönülten helyezkednek el a flow cell felületén. Ez a szekvenálás során jelentkező fényjelek detektálásánál jelentős, hiszen egy klaszter által generált fényjelzés jóval erősebb, mint egy egyedi molekulából érkező felvillanás. Az elv emellett növeli a szekvenálás statisztikai bizonyosságát: ha egy klaszter egy elemében beépül egy hibás nukleotid, a hibás beépülés fényjelét elnyomja a klaszter többi eleméből származó helyes felvillanás, a szekvenátor kamerája pedig ezt azonosítja majd.



2. ábra: A rövid leolvasású Illumina szekvenátor működési elve

A klasztergenerációt a szekvenálás követi. A flow cellre primereket, DNS-polimeráz enzimet, illetve módosított dezoxiribonukleotid-trifoszfátokat (dNTP-k) viszünk fel. A módosított dNTP-k beépülése fluoreszcens fényfelvillanással jár, mely a négy különböző nukleotid esetén eltérő színű. A szekvenátorban kamera található, mely minden beépülési esemény alatt lefényképezi a flow cellt. Az egyes klaszterek helyén tapasztalt színjelenség alapján a szekvenátor beépített számítógép segítségével azonosítja az egyes molekulákba az egyes szintetikus események során beépült nukleotidokat, a vizsgált nukleinsavak szekvenciája pedig így megállapítható. Ez a folyamat a bázishívás vagy basecalling. A bázishívás outputja a betűkkel jelölt nukleotidsorrend fastq fájlformátumban.

A nukleotidok módosításai miatt a polimeráz egyszerre csak egy nukleotidot képes a szintetizálódó szála építeni, a módosítást a nukleotidbeépülés után kémiai úton távolítja el a rendszer. Ezután a be nem épült dNTP-k lemosódnak a flow cellról, és újak érkeznek, azaz új beépülési ciklus zajlik le. A beépülési és lemosási lépések ciklusai a flow cellhoz rögzítve szekvenált DNS-molekulák komplementer szálainak teljes megszintetizálásáig zajlanak.

A szintetikus szekvenálási események a Sanger-szekvenálással szemben párhuzamosan történnek: egy flow cellen akár 96 minta leolvasása történhet egyszerre. Széklet metagenom szekvenálás esetében ez azt jelenti, hogy egy szekvenálás során legfeljebb 96 különböző alanytól és/vagy eltérő időpontban vett minta teljes DNS-tartalma szekvenálható. Az első mintához a DNS-izoláció során az 1. adaptert ligáljuk, a másodikhoz a 2. adaptert, és így tovább, akár 96 különböző széklet izolátum eredetű metagenom különböztethető meg. Az egyes mintákban található DNS-molekulák nagy számát, illetve az azokból generált klasztereket véve, egy szekvenálás során akár tízmillió egyedi szekvencialeolvasás is történhet az Illumina MiSeq szekvenátoron.

A szekvenálás lehet paired-end (3. ábra) és single-end. A single-end leolvasások a szekvenált fragmentum egy oldaláról történnek, paired-end leolvasások esetében viszont mindkét oldalról. A hosszabb szekvenciák leolvasására a paired-end megközelítés kedvezőbb, mivel a két végéről leolvasott szekvencia alapján létrejött readek bioinformatikai módszerekkel összeilleszthető és a teljes szekvencia visszakapható abban az esetben is, ha annak hossza meghaladja a szekvenátor leolvasási kapacitását.



3. ábra: Paired-end szekvenálás

4.3 Célzott amplikon szekvenálás és a riboszomális V régiók

Amplikon szekvenálással célzott génrégiókat vizsgálunk. A teljes DNS-t az izolációt követően nem fragmentáljuk hasonló méretű darabokra szonikálással vagy transzpozáz enzim segítségével, hanem abból célzott PCR-amplifikációval felszorzozunk egy kívánt szakaszt, a szekvenálási könyvtárat pedig a létrejött amplikonokból készítjük.

A prokarióta riboszóma kis alegység 16S rRNS-ének génje hiperkonzervatív és hipervariábilis szakaszokat is tartalmaz, utóbbiak taxonómiai markerekként használhatók. A hiperkonzervatív régiókra tervezett univerzális primerekkel tetszőleges variábilis, azaz V régiók amplifikálhatók, ezek bázissorrendje szekvenálást követően bioinformatikai módszerekkel elemezhető.

Kutatócsoportunk a 16S rRNS gén három hipervariábilis régióját hordozó V1-V3, illetve a V4 régió amplikonjait elemezte. A V1-V3 amplikonok teljes hossza körülbelül 600 bázispár. Az alkalmazott Illumina MiSeq szekvenátor rövid leolvasásokat generál, így egy 600 bázispár hosszú szakasz terminális bázisainak minősége paired-end megközelítéssel is alacsony lehet. Ezen technikai limitáció következtében a V1-V3 amplikonok középső szakasza - azaz feltehetően a V2 régió terminális és az azt követő konzervatív régió iniciális szegmense - kis megbízhatósággal lett leolvasva, illetve később, a bioinformatikai elemzés során, eltávolításra került. Ezek ellenére a szekvencia primerekhez proximális, nagy hűséggel leolvasott V1 és V3 régiói elméletileg megteremtik a taxonómiai azonosításhoz szükséges követelményeket. Az összesen körülbelül 120 bázispár hosszú V4 amplikonok esetében ez a probléma nem állt fent.

4.4 A 16S rRNS amplikon szekvenálási adatok kiértékelése bioinformatikai módszerekkel

Az Illumina MiSeq szekvenátorba beépített számítógép saját programja segítségével fordítja le az egyes dNTP-k beépülésekor tapasztalt felvillanásokat a megfelelő bázist jelző betűre. Az output az egyes leolvasások szekvenciáját együtt tartalmazó fastq fájl, ezen adatokat számítógépes programok segítségével tudjuk szűrni és taxonómiai abundanciaértékek vagy más vizsgálni kívánt jellemzők tekintetében elemezni. A szekvenált minták a bioinformatikai elemzés során a minták egyedi azonosító barcode-ja alapján különböztethetők meg egymástól.

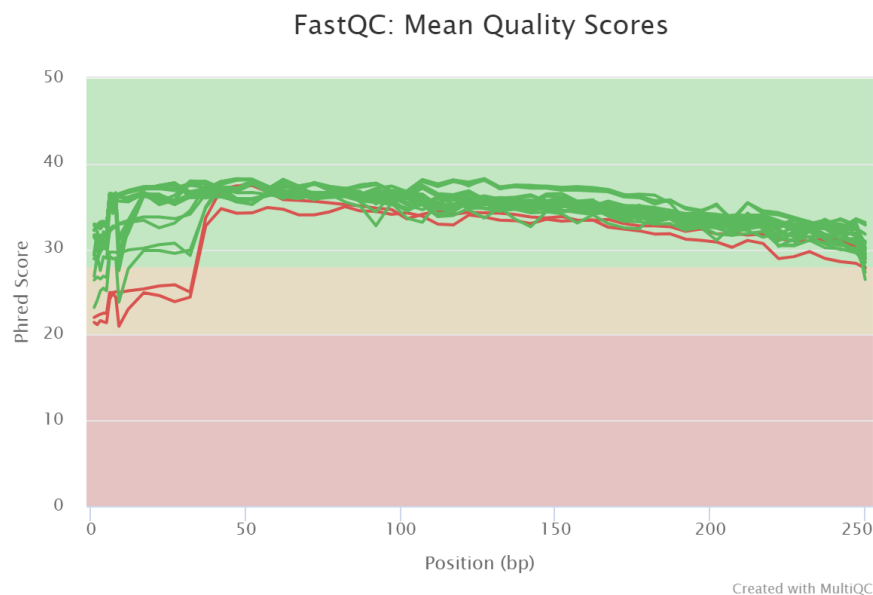
Munkacsoportunk a DADA2 R alapú programcsomagot (Callahan, McMurdie, Holmes 2017) alkalmazta a szekvenciák elemzésére, mely az RStudio fejlesztői környezetben fut. A DADA2 más, korábbi fejlesztésű taxonómiai azonosítást végző programokhoz képest komoly módszertani előnnyel bír.

A korábbi szoftverek az operációs taxonómiai egységek, azaz OTU-k megkülönböztetésén alapulnak. Az OTU a taxon fogalmának megfelelő jelentéssel bír a molekuláris taxonómiában: a kifejezés egy választott marker gén szekvenciája alapján egymáshoz rendelt organizmusok csoportját jelöli. Az OTU-k megállapítására kétféle mód van. A *de novo* OTU-k esetében egy bizonyos, általában 97%-os konszenzus szekvencia egyezési limitnél nagyobb egyezést mutató szekvenciák alkotnak egy OTU-t, a *closed reference* OTU-k megállapításához pedig a vizsgált szekvenciákat egy referencia adatbázishoz hasonlítjuk. Ha az adatbázis egy elemének a vizsgált szekvencia megfeleltethető, akkor azt az adott referencia taxontól származónak tudjuk be, és minden ezzel egyező szekvenciát ugyanabba az OTU-ba sorolunk. Összegezve tehát, ha két szekvencia egymással megfelelő mértékű egyezést mutat, vagy megfelel egy referencia adatbázis

egy elemének, akkor ugyanabba az operációs taxonómiai egységbe kerül. Az OTU alapú módszerek tehát több egyedi szekvenciából álló klasztereken végeznek műveleteket.

A DADA2 nem OTU-kat, hanem ASV-ket, azaz amplikon szekvencia variánsokat alkalmaz (Callahan, McMurdie, Holmes 2017). Az ASV-k a célzott amplikon szekvenálás során létrejött egyedi leolvasások: a program felismer minden egyes megszekvenált 16S rRNS gén amplikont, és megállapítja, hogy a mintában hányszor fordulnak elő az egyes leolvasások. Ez elengedhetetlen a jelenlévő taxonok abundanciaértékének megállapításához.

Az ASV-k azonosításának kezdeti lépéseként a program egy hibamodellt futtat a szekvenálási adatsoron, mellyel megállapítja, hogy a nagy hasonlóságú leolvasások különbségei biológiai jellemzők vagy szekvenálási hiba folytán jelentek meg. A hibamodell-futtatást követően nagy bizonyossággal állítható, hogy az akár igen kismértékű különbségeket prezentáló szekvenciák egymástól valóban eltérnek, miáltal nagy pontosságú fajsztípus vagy akár faj alatti (baktériumtörzsek, *strain*-ek) taxonómiai azonosítás lehetséges. Az OTU alapú módszereknél gyakori hiba, hogy a közeli rokon fajokat a program nem különbözteti meg, az ASV-alapú programokra ez azonban nem jellemző. A hibamodellnek hála a fals pozitív eredmények aránya is kisebb.



4. ábra: A leolvasott bázisok minőség vagy quality értékének vizuális reprezentációja. A quality értékek a 250. leolvasott bázis környékén csökkennek

A hibamodell lefuttatását egy szűrési lépés követi. Az Illumina szekvenálás rövid leolvasású szekvenálás, egy bizonyos leolvasási hossz után a leolvasott bázisok megbízhatóságának nagymértékű csökkenése tapasztalható (4. ábra). A szűrés során a szekvenciavariánsok egy megadott alacsonyabb minőségi értéket mutató terminális szakasza, illetve az adapter szekvencia kiselejtezésre kerül, aminek következtében a továbbiakban vizsgált, szűrt ASV-k szekvenciája ismert és definiált statisztikai bizonyosságot mutat.

A szűrési vagy filtering lépés után merge, azaz egyesítési lépés történik. A paired-end leolvasással ugyanazon amplikonról készült ellentétes lefutású readeket az adapterük szekvenciája alapján illeszti össze a program.

Az egyesítést követően az ASV-k taxonómiai azonosítása referencia adatbázis segítségével zajlik. Kutatócsoportunk a SILVA (Quast és mtsai. 2012) 16S rRNS online génbankot alkalmazta referenciaként. A DADA2 program nagy előnye, hogy a beazonosítatlan leolvasások ugyanúgy megmaradnak az outputban, mely lehetőséget ad a vizsgált közösségek korábban le nem írt tagjainak azonosítására.

A DADA2-vel azonosított taxonok abundanciaértékeit, a különböző egyedektől különböző időpontokban gyűjtött minták hasonlóságait és különbségeit, illetve egyéb mutatók vizuális megjelenítésére további programok - például a csoportunk által használt Phyloseq (McMurdie, Holmes 2013) - szükségesek, a kutatási eredmények értékelése során szükség van az ilyen jellegű adat-interpretációkra. Jelen kutatás során is nagyban támaszkodtunk ezekre, amint igyekeztünk megállapítani, hogy a vizsgált 16S rRNS gén V régió amplikon szekvenálási módszerek közül melyik alkalmasabb intesztinális mikrobaközösségek vizsgálatára.

Célkitűzés

Jelen pályamunka célja összehasonlítani két metagenom kompozíciós elemzési módszert 20 kutya székletminta metagenomikai elemzése alapján. Az alkalmazott módszerek a prokarióta 16S rRNS gén V1-V3 és V4 hipervariábilis régiók amplikonjainak szekvenálása, és a szekvenciák azonosítása a DADA2 program és a SILVA 16S rRNS adatbázis alapján. A kiemelten vizsgált minták 6 állattól, állatonként 3 vagy 4, összesen 7 mintavételi időpontból származnak. A mintázott állatok egy felnőtt hím, egy felnőtt nőstény és a nőstény négy kölyke. A felnőtt egyedektől a kölykök születése előtt és után is gyűjtöttünk mintát, a kölykök mintáit az első fél életév során gyűjtöttük.

Az elemzés során mindkét módszerrel elvégeztük a kölyök és felnőtt állatok mikrobiomjának longitudinális elemzését, különös tekintettel a kölyök mikrobiomok korai fejlődésének jelenségeire. A két módszerrel kapott eredményeket célunk volt összehasonlítani, megállapítani az eredmények metszeteit és különbségeit, végső soron pedig azt, hogy a két alkalmazott módszer közül melyik ad részletesebb, nagyobb felbontású képet a vizsgált mikrobiomokról.

Mivel további kutatásaink során nagyszámú kutya és humán eredetű minta feldolgozását tervezzük, fontosnak tartjuk a munkafolyamatok korai szakaszában értékelni az alkalmazandó módszereket. A széklet metagenomi DNS vizsgálata során a kiválasztott módszer, például a szekvenálásra választott riboszomális V régió, komolyan befolyásolhatja a kapott eredményeket, a nem megfelelő módszer jelentette torzulás pedig kedvezőtlenül befolyásolhatja a kutatások értékét.

Mikrobiomikai kutatómunkánk távlati céljai közé tartozik körülbelül 200, elhízásra és társbetegségeire nézve eltérő mutatókkal bíró, különböző diétás-és életmódprogramokat követő humán páciens bélflórájának követéses vizsgálata, illetve 100-120, különböző szénhidráttartalmú tápon nevelt, többségében fiatal kutya modellállat mikrobiomjának longitudinális elemzése. Vizsgálataink során többféle mikrobiom, virom és gazda genom szekvenálási eljárást alkalmazva igyekszünk feltárni az étrend, a mikrobiom és az egészség komplex kapcsolatát. Eredményeink új diagnosztikai és terápiás eljárások alapjait fektethetik le, kutya mikrobiomikai eredményeink kisállat-egészségügyi jelentőséggel is bírhatnak.

Anyagok és módszerek

1. Modellállatok és mintagyűjtés

Mérlegelve a mikrobiomikai kutatás során alkalmazott egyes modellszervezetek előnyeit és hátrányait, kutatócsoportunk választása a kutyára (*Canis lupus familiaris*) esett, melynek mikrobiomja a humán bélflórával nagy hasonlóságot mutat (Coelho és mtsai. 2018). Fajtatiszta és ismert táplálású kutyákat mintázunk, a Magyar Pumi Klub tenyésztőivel együttműködve számos magyarországi kennel pumikutyájától kapunk nukleinsav-izolálásra alkalmas székletmintát. A születésüktől fogva mintázott kutyák esetén a minták gyűjtése, illetve a táplálék dokumentációja az új gazdához kerülés után is zajlik. Ezáltal számos különböző korú, rokonsági viszonyú és étrendű állat longitudinális mikrobiom-elemzését végezzük, mely alapján a mikrobiom és az étrend, illetve a genetika kapcsolatáról értékes adatok gyűjthetők.

1.1 Vizsgálati alanyok és étrendjük

A Duna-menti Dumás Kennel egyike tenyésztő partnereinknek. Jelen pályamunkában a DMD 2020/B almának 4 kölyökkutyájától, a kölykök anyjától és a tenyészethez tartozó, de a vizsgált egyedekkel közvetlen rokonságban nem álló kontroll hím egyedtől gyűjtöttünk mintát fél éven keresztül, összesen 7, egyedenként 3 vagy 4 mintavételi időpontban.

A vizsgálat időtartama alatt a felnőtt egyedek (Suba¹ anyakutya, a továbbiakban A; Lidérc kontroll hím, továbbiakban Fk) ugyanazon tápot² kapták (Brit Care Dog Show Champion Salmon and Rice), a kölykök étrendje a kennelben tartózkodás ideje alatt volt egységes (Brit Care Grain Free Puppy Salmon and Potato), az új gazdához kerülést követően változott. A kölykök a 4. héttől kaptak először tápot, a hozzátáplálás a teljes átállásig növekvő gyakorisággal és mennyiségben zajlott. Az egyes állatok esetében az említett tápok tették ki az egyedek kalóriabevitelének zömét, ám kis mennyiségben a jelzett táptól eltérő tápanyagok bevitel is előfordulhatott, például a tréningezés során adott jutalomfalatoké. A tápok részletes összetétele az 1. függelékben található. A kölykök étrendjére és új élőhelyére vonatkozó információkat a következőkben részletezzük:

¹ Az állatokra a törzskönyvezett nevük helyett egyedi kóddal fogok hivatkozni

² A tápok analitikai összetétele a függelékben található

- Boszi (K1) a vizsgálat során végig a tenyésztőnél maradt, a hozzátáplálás kezdetétől (4. hét) a vizsgálat végéig a fent említett lazacos-burgonyás tápot kapta.
- Blöff (K2) a hozzátáplálás kezdetétől (4. hét) a vizsgálat végéig a fent említett lazacos-burgonyás tápot kapta. 2 hónaposan (2021 január 3-án) új gazdához került egy Pest megyei kertés házba, elsősorban benti kutyaként.
- Borcsa (K3) a hozzátáplálás kezdetétől (4. hét) 2 hónapos koráig (2021 január 3-ig) a fent említett lazacos-burgonyás tápot kapta; 2021 január 3-án új gazdához került egy fővárosi lakásba, ahol a Brit Care márka egy eltérő összetételű kölyökkutya-tápjával (Brit Care Puppy All Breed Lamb and Rice) etették a vizsgálat végéig.
- Briós (K4) a hozzátáplálás kezdetétől (4. hét) 2,5 hónapos koráig (2021 január 20-ig) a korábban említett lazacos-burgonyás tápot kapta; 2021 január 20-án, négy nappal a 10. heti mintavételt követően új gazdához került egy svájci lakásba, ahol a BARF protokoll szerint etették a vizsgálat végéig. A BARF (Bones and Raw Food/Biologically Appropriate Raw Food) étrend a kutyák természetes táplálkozását imitáló nyers étrend. A K4 kölyök étrendje a vizsgálat időtartama alatt 75%-ban nyers húsból és 25%-ban nyers zöldségből és gyümölcsből (ld. 1. függelék) állt. A BARF diéta alacsony szénhidráttartalmú vagy low-carb diétának számít, a napi bevitt kalóriák 3-6 százaléka szénhidrát.

1.2 Mintavételezési időpontok, minták kezelése

A vizsgált állatoktól tenyésztőik vagy gazdáik előre egyeztetett időpontokban gyűjtöttek székletmintát. A minták gyűjtése az ürítéstől számított 15 percen belül megtörtént, majd átadásig -20, laboratóriumunkba kerülve -80 Celsius fokon voltak tárolva. Különleges esetekben, mint a külföldi gazdához került K4 egyed esetében a minták szobahőmérsékleten, speciális tárolókban, a nukleinsavak integritását biztosító folyadékban (ZymoResearch DNA/RNA Shield) érkeztek laboratóriumunkba.

Az egyes egyedekre vonatkozó mintavételi időpontokat a következőkben listázzuk (*1. táblázat*). A számok a kiskutyák születése óta eltelt időt jelentik napban, a felnőtt egyedeknél látható -7 a szülés előtt 7 nappal történt mintavételre utal.

A kölyköktől az első hetekben nehézkes a mintavétel, ugyanis az anyaállat a fiatal utódok ürülékét gyakran elfogyasztja. Jelen pályamunkában 5 hetesnél korábbi minta nem került feldolgozásra, a

későbbi kutatás során viszont hangsúlyt kap a korai minták elemzése, ugyanis a bélflóra legkorábbi összetétel-adatainak figyelmen kívül hagyása a kutatás értékét csökkentené.

Bizonyos állatoktól a vártnál több alkalommal kaptunk mintát. Minden mintát feldolgoztunk és részletesen kielemeztünk, ám az eredmények értékelése során azonos mintavételi időpontokra igyekeztünk fókuszálni, így pályamunkánk során kizárólag a kitüntetett időpontokban vett mintákról értekezünk. A kitüntetett időpontok félkövérrel vannak jelölve.

1-V3 amplikon szekvenálás	4 amplikon szekvenálás
-7, 1, 180	-7, 1, 180
-7, 1, 180	-7, 1, 180
1: 37, 70, 180	1: 37, 70, 180
2: 37, 39, 70, 180	2: 37, 39, 70, 180
3: 37, 70, 180	3: 37, 70, 180
4: 37, 51, 70, 180	4: 37, 51, 70, 180

1. táblázat: Mintavételi időpontok

A felnőtt egyedektől a kölykök születése előtt 7, illetve utána 1 nappal és 180 nappal (6 hónappal) került elemzésre minta, a kölyköktől 37 napos (5 hetes), 70 napos (10 hetes) és 180 napos (6 hónapos) mintákat dolgoztunk fel, kivéve K2 esetében. Itt a 37. napi mintát nem dolgoztuk fel V4 amplikon szekvenálással. A 39. napi minta mindkét szekvenálási eljárással feldolgozásra került, így K2 korai mintáját ebben az időpontban vizsgáljuk.

A kölykök 5 hetes korukban anyatejet, illetve hozzátáplált kölyök tápot kaptak. 10 hetesen és 6 hónaposan étrendjük a fentebb részletezettek szerint divergált.

2. DNS izolálás, könyvtárkészítés és szekvenálás

2.1 DNS-izolálás kutya székletmintából

A 16S rRNS gén V1-V3 és V4 régióira célzott amplikon szekvenálás könyvtárát a velünk együttműködő tenyésztők és gazdák által gyűjtött, laboratóriumunkban -80 Celsius fokon tárolt pumi székletmintákból izolált DNS-ből készítettük el. A könyvtárkészítés során teljes széklet metagenomból kell a kívánt V régiókat amplifikálni, az amplikonokat pedig el kell látni a megfelelő szekvenálási adapterekkel. Ezen munkafolyamatokat kereskedelmi forgalomban kapható kitékkel végeztük.

A -80 Celsius fokról kivett mintákból harminc perces, szobahőmérsékleten, elszívófülke alatt történő melegítés után 0,32-0,38 grammos darabokat vágunk le, a darabokból a QIAAmp Fast Stool DNA Mini Kit (50) - Pathogen Protocol kit protokollja szerint 100 mikroliteres végtérfogatban izoláltunk teljes metagenomi DNS-t.

2.2 Amplikon szekvenálási könyvtárkészítés

A 100 mikroliteres mintákból a NEXTflex 16S V1-V3 Amplicon-Seq Kit és a NEXTflex 16S V4 Amplicon-Seq Kit kitékkel készítettünk V1-V3, illetve V4 amplikon szekvenálási könyvtárakat 16 mikroliter végtérfogatban.

Az amplikon szekvenálási könyvtár-készítési protokollok során két egymást követő PCR-amplifikáció zajlik. Az első PCR során történik a széklet metagenomi DNS-ből a megfelelő konzervatív régiókra specifikus univerzális primerek segítségével a V1-V3, illetve V4 régiók felsokszorozása, a második PCR során pedig az előző lépés amplikonjait amplifikáljuk olyan primerekkel, melyek tartalmazzák a megfelelő Illumina szekvenálási adaptereket. A második PCR lépés eredménye a MiSeq-kel megszekvenálható 16S rRNS V1-V3, illetve V4 könyvtár, könyvtáranként 16 mikroliter végtérfogatban.

2.3 16S rRNS V1-V3 és V4 amplikon könyvtárak szekvenálása

A könyvtárak szekvenálását az Illumina MiSeq platform protokollja szerint végeztük. Mind a V1-V3, mind a V4 amplikonok szekvenálásakor a MiSeq Reagent Kit v2-t használtuk, a ciklusszám 500 volt.

3. A szekvenálási adatok bioinformatikai és manuális elemzése

3.1 Amplikon szekvencia variánsok taxonómiai azonosítása, az adatok vizualizációja

A bázishívást (basecalling, a szekvenátor által detektált egyedi fényjelenségek nukleinsav-bázisokra fordítása) az Illumina MiSeq saját szoftvere végezte. A keletkezett fastq formátumú fájlokat az amplikon szekvencia variáns, azaz ASV-alapú DADA2 1.20.0 verziójával, RStudioban dolgoztuk fel. A DADA2 munkafolyamat lépései a hasonló szekvenciák különbségeit elemző *hibamodell-futtatás*, az alacsony quality értékű bázisok és az adapterszekvenciák eltávolítását jelentő *szűrés*, és a paired-end szekvenálás során keletkező komplementer leolvasások egyesítése, azaz *merge*.

A vizsgált amplikon szekvenciák szűrése során 245 bázispáros szűrési értékkel dolgoztunk. A V1-V3 amplikonok teljes hossza körülbelül 600 bázispár, megszekvenálásuk paired-end leolvasással történt. A beállított szűrési érték miatt az egyesített amplikon szekvencia középső szakaszáról egy körülbelül 100 bázispárnyi szakasz elvész. Az elvesztett szakasz feltehetően a V2 variábilis régió terminális és az azt követő konzervatív régió iniciális része, mindazonáltal a teljes kiterjedésében ismert V1 és V3, illetve a részlegesen ismert V2 szekvenciák alapján fennállnak az ASV-k azonosításához szükséges követelmények. Az ennél rövidebb V4 amplikonok esetében ez a probléma nem lépett fel. A V1-V3 merge-ek hossza 490, a V4 merge-eké körülbelül 250 bázispár.

A folyamat végére a programmal nagy bizonyossággal megkülönböztetett ASV-eket a 16S rRNS referencia adatbázishoz (SILVA version 138) hasonlítva azonosítottuk taxonómiailag. A szekvenciák taxonómiai azonosítása nemzetség szintig volt pontos és megbízható. A nehezen kezelhető, táblázat formátumú adatok vizuális reprezentációját a Phyloseq 1.36.0 verziójával végeztük, szintén RStudio fejlesztői környezetben. A DADA2-vel feldolgozott és Phyloseq-kel vizualizált adatokat HTML-fájlként mentettük. A fájlok tartalmazzák az egyes minták 20 legabundánsabb taxonját bemutató oszlopdiagramokat törzs, osztály, rend, család, nemzetség és - adott esetben - faj szinteken, az egyes minták Shannon és Simpson index szerinti diverzitását, a minták hasonlóságait bemutató klaszter dendrogramot, a minták főkomponens-analízisét és nonmetrikus többdimenziós skáláját. A fenti paramétereket így mind a V1-V3, mind a V4 amplikon szekvenálással feldolgozott mintákra megkaptuk.

3.2 Manuális adatelemzés

A manuális elemzés során megvizsgáltuk, hogy az egyes módszerekkel az összes mintára vetítve hány ASV-t sikerült azonosítani, illetve minden faj feletti taxonómiai szinten megállapítottuk, hogy az adott szinten vizsgált 20 leggyakoribb taxon közül hányat sikerült mindkét módszerrel azonosítani. Azért fókuszáltunk a leggyakoribb taxonokra, mivel a ritkébbak abundanciája igen kicsi, szerepük a gyakori taxonoknál feltehetően jóval kevésbé jelentős, emellett a módszerek összehasonlításához ezek az adatok mérvadóak.

Minden mintában és minden taxonómiai szinten megállapítottuk az egyes mintákban és egyes rendszertani szinteken fellelhető taxonokat, és igyekeztük a vizsgált minták összetételében mintázatokat felismerni. Az elemzés során a minták összetételéről részletes áttekintést kaptunk.

Minden faj feletti taxonómiai szinten megállapítottuk, hogy az adott szinten vizsgált 20 leggyakoribb taxon közül hányat sikerült mindkét módszerrel azonosítani.

Elvégeztük a minták részletes összehasonlítását, elsősorban törzs és nemzetség szinten. Egy bakteriális phylum igen eltérő funkciójú tagokat számlálhat, ám a törzsek vizsgálata általános tendenciák felismerését lehetővé teszi. Adott törzsek jelenléte és relatív gyakorisága durván utalhat bizonyos metabolikus funkciókra, egyes törzsek és egyes folyamatok között kimutatható asszociáció. A nemzetségek szintjén történő elemzés azonban nagyobb felbontású betekintést nyújt a bélben zajló folyamatokra. Habár genuson belül is igen eltérő életműködésű mikrobák fordulhatnak elő, az egy nemzetség által betöltött ökológiai és metabolikus funkciók jóval behatárolhatóbbak. A mintákban fellelhető törzsek és nemzetségek szintjén elvégzett összehasonlítása után magyarázatokat kerestünk a látható hasonlóságokra és különbségekre. Ennek során a mintavételezési időponthoz köthető adatokra (kutya kora, tápláléka, lakóhelye), valamint irodalmi referenciákra támaszkodtunk. Ezen adatokat kiterjedésüknél fogva részletesen nem közöljük jelen pályamunkában, azonban bemutatunk bizonyos eseteket, melyekben a két módszerrel kapott adatok kiemelkedő mértékben eltérnek.

A két amplikon szekvenálási analízis által kiadott klaszter dendrogramokat, azaz a minták egymáshoz való hasonlóságát reprezentáló ábrákat, összehasonlítottuk. Megvizsgáltuk, hogy a két módszer milyen eltéréseket prezentál a minták hasonlósági viszonyait tekintve. A minták diverzitásértékeit is összehasonlítottuk a két módszer szerint.

Értékeltek a két amplikon szekvenálási megközelítés előnyeit és hátrányait, megállapítottuk munkánk lehetséges limitációit és további rövidtávú kutatási célokat fogalmaztunk meg.

4. A munkafolyamat lépései összefoglalva

Minták gyűjtése (tenyésztő vagy gazda), minták beérkezése a laboratóriumba

tárolás hűtve vagy ZymoResearch DNA/RNA Shield-ben

DNS izolálás - QIAamp Fast Stool DNA Mini Kit (50) - Pathogen Protocol

Amplikon szekvenálási könyvtár készítés

Perkin Elmer NEXTflex 16S V1-V3 Amplicon-Seq Kit 12

16S rRNS gén V1-V3 régió PCR amplifikációja

adapterek ligálása az amplikonokra PCR segítségével

Perkin Elmer NEXTflex 16S V4 Amplicon-Seq Kit 12

16S rRNS gén V4 régió PCR amplifikációja

adapterek ligálása az amplikonokra PCR segítségével

Rövid leolvasású paired-end amplikon szekvenálás - Illumina MiSeq

a V1-V3 és V4 amplikon szekvenálások során alkalmazott szekvenálási kémia:
MiSeq Reagent Kit v2

a V1-V3 és V4 amplikon szekvenálási ciklusok száma: 500

Bioinformatikai adatelemzés - DADA2 v.12.20.0

Hibamodell futtatása

Szekvenciák szűrése

Szekvenciák egyesítése

ASV-k SILVA v.138 referencia adatbázishoz hasonlítása

Adatok vizuális megjelenítése - Phyloseq v.1.36.0

Adatok manuális elemzése

Eredmények és értékelésük

Munkánk során a humán mikrobiom modellorganizmusaként alkalmazott fajtatiszta pumikutyák mikrobiomját vizsgáltuk két 16S rRNS gén V régió amplikon szekvenálási megközelítéssel. 6 állat kitüntetett mintavételi időpontokban gyűjtött székletmintájából izolált metagenomi DNS-sel dolgoztunk.

1. Azonosított ASV-k

Eredményeink elemzése során bizonyos eltéréseket tapasztaltunk a V1-V3 és a V4 amplikon szekvenálással azonosított taxonok között. Számos taxont mindkét módszerrel kimutattuk a mintákban, ám egyes ASV-k rendre csak a V1-V3 vagy csak a V4 megközelítést alkalmazva jelentek meg.

V1-V3 amplikon szekvenálással összesen 20, V4 amplikon szekvenálással összesen 19 mintát vizsgáltunk. Az egy adott módszerrel vizsgált összes mintát tekintve, a V1-V3 amplikon szekvenciák alapján összesen 2210 amplikon szekvencia variánst, ASV-t azonosítottunk. Az összes ASV azonosítása után a 0,05%-nál nagyobb abundanciájú ASV-kre szűrtünk: a V1-V3 megközelítés alapján 195 ASV volt jelen fél ezreléknél nagyobb gyakorisággal az összesen vizsgált 20 mintában. V4 amplikon szekvenálással 19 mintából összesen 1235 ASV-t tudtunk azonosítani, melyek közül 235 ASV gyakorisága volt 0,05% fölötti.

Minden faj feletti taxonómiai szinten megállapítottuk, hogy az adott szinten vizsgált 20 leggyakoribb taxon közül hányat sikerült mindkét módszerrel azonosítani (2. táblázat). Azért fókuszáltunk a leggyakoribb taxonokra, mivel a ritkébbak abundanciája igen kicsi, szerepük a gyakori taxonoknál feltehetően jóval kevésbé jelentős. Eredményeink alapján 11 törzset, 14 osztályt, 20 rendet, 17 családot és 10 nemzetséget mind a V1-V3, mind a V4 amplikonok szekvenálása alapján ki tudtunk mutatni a vizsgált mintákból.

A felsorolt taxonok mellett mindkét módszerrel sikerült olyan taxonokat azonosítanunk, melyeket a másik megközelítés nem mutatott ki: V1-V3 szekvenálással további 1 törzset, 3 osztályt, 2 rendet, 3 családot és 8 nemzetséget, V4 megközelítéssel további 5 törzset, 6 osztályt, 2 rendet, 1 családot és 6 nemzetséget azonosítottunk az egyes szinteken leggyakoribb 20 taxon közül - azaz a V4 a rend feletti, a V1-V3 módszer pedig a rend alatti taxonómiai szinteken azonosított több csoportot, a referencia irodalom adataival egyetértésben (Bukin és mtsai. 2019). Ezek mellett egyes taxonómiai szinteken beazonosíthatatlan taxonokat is detektáltunk, mivel a DADA2 program a referencia adatbázis alapján nem azonosított ASV-eket is megtartja az outputban. Így V1-V3 szekvenálással

osztály és nemzetség, V4 szekvenálással család és nemzetség szinteken be nem azonosított taxonokat detektáltunk. Ezen ASV-k esetében a taxonómiai azonosítás csupán magasabb rendszertani kategóriák szintjén volt sikeres.

azonosított taxonok	V1-V3 ampikon-seq	V4 ampikon-seq	mindkét módszer
faj	2	5	
osztály	7	0	4
rend	0	0	3
család	0	3	7
nemzetség	3	5	0

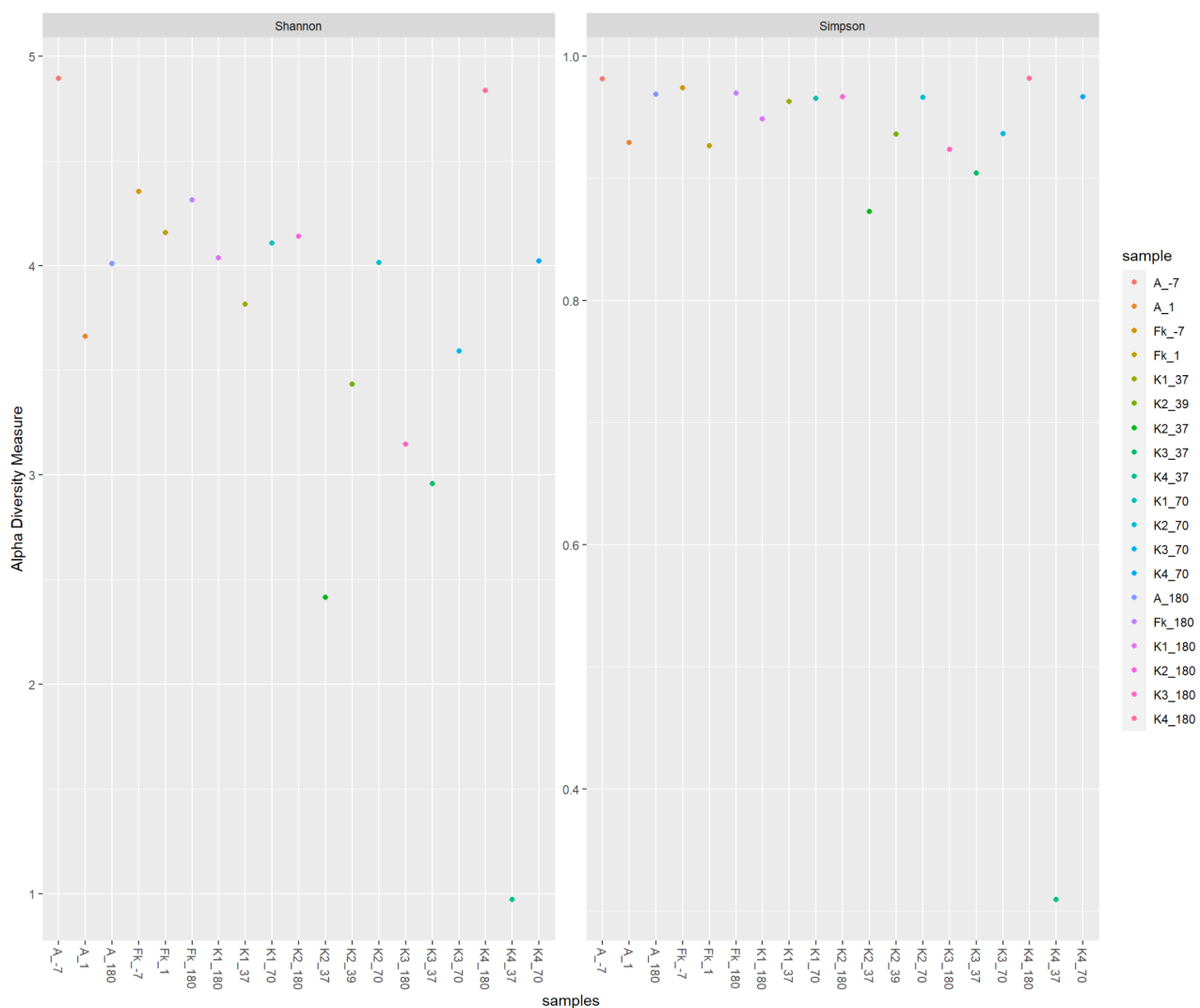
2. táblázat: Az egyes ampikon szekvenálási módszerrel azonosított taxonok száma az egyes faj feletti taxonómiai szinteken, a minták 20 leggyakoribb taxonjain belül

A V1-V3 és V4 megközelítésekkel külön-külön, illetve átfedésben azonosított taxonok listáját a Függelék tartalmazza.

A fajszerű azonosítás mindkét módszerrel korlátozott eredménnyel járt: V1-V3 szekvenálással összesen 7, V4 szekvenálással összesen 5 faj jelenlétét sikerült nagy bizonyossággal megállapítani, így a fajszerű elemzést jelen munkánkban nem részletezzük.

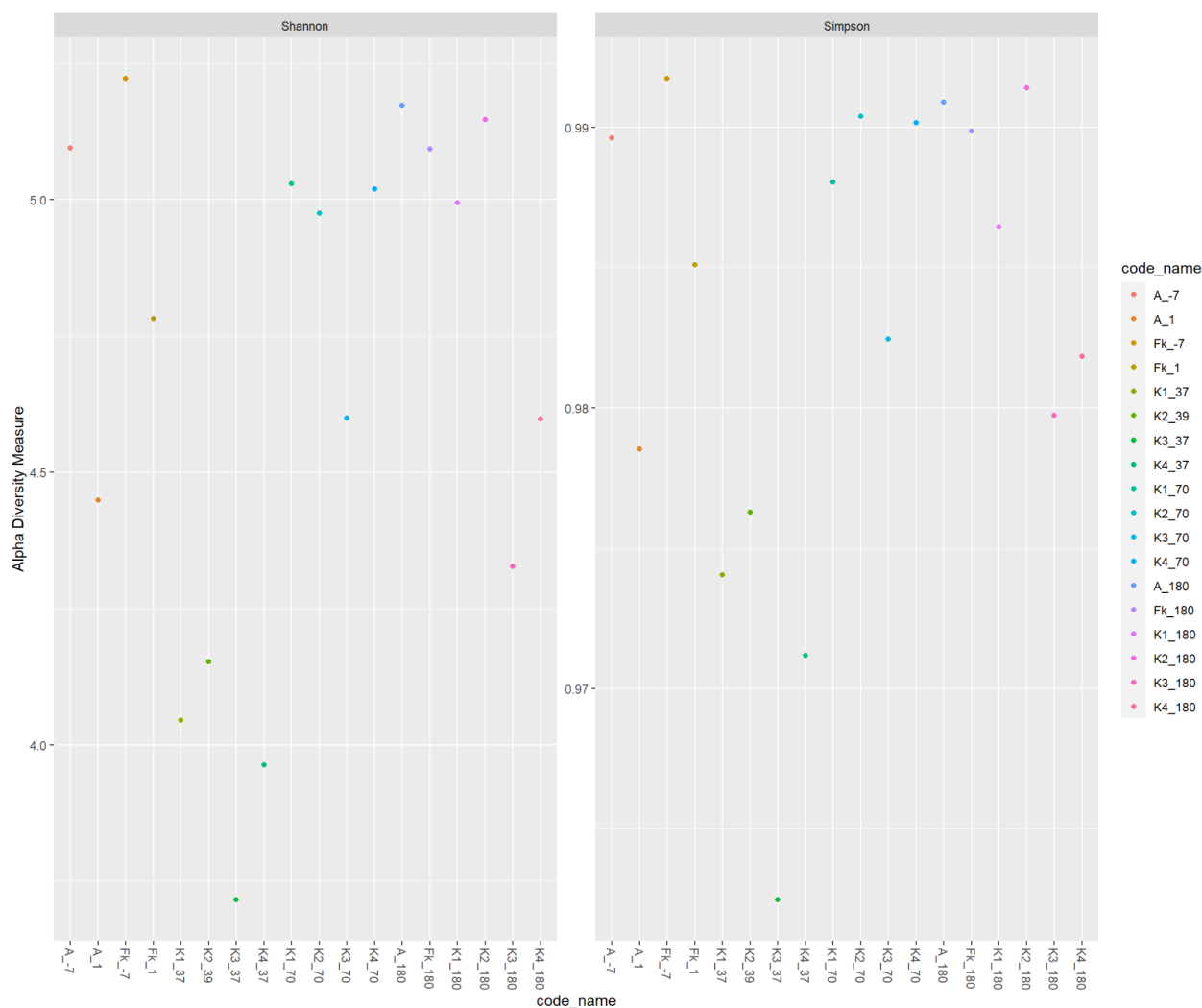
2. Diverzitásindexek

A két vizsgálati módszer eltérő eredményeket hozott a vizsgált minták diverzitási értékeit tekintve. A minták alfa diverzitását Shannon és Simpson indexek szerint is vizsgáltuk. A diverzitásértékeket a kitüntetett időpontokban vett összesen 18 mintára vetítettük.



5. ábra: V1-V3 diverzitási indexek

A V1-V3 szekvenálással vizsgált 18 minta (5. ábra) közül 15 mutatott legalább 3-as értéket a Shannon indexen, 11 minta értéke 4 feletti. 2 minta diverzitása 2 és 3 feletti, egyé 1 alatti. A Simpson index szerint 18-ból 16 minta 0,9 feletti diverzitásértéket mutat. A két további minta egyike megközelíti a 0,9-es értéket, a másik alacsony, 0,3 közeli diverzitást mutat.



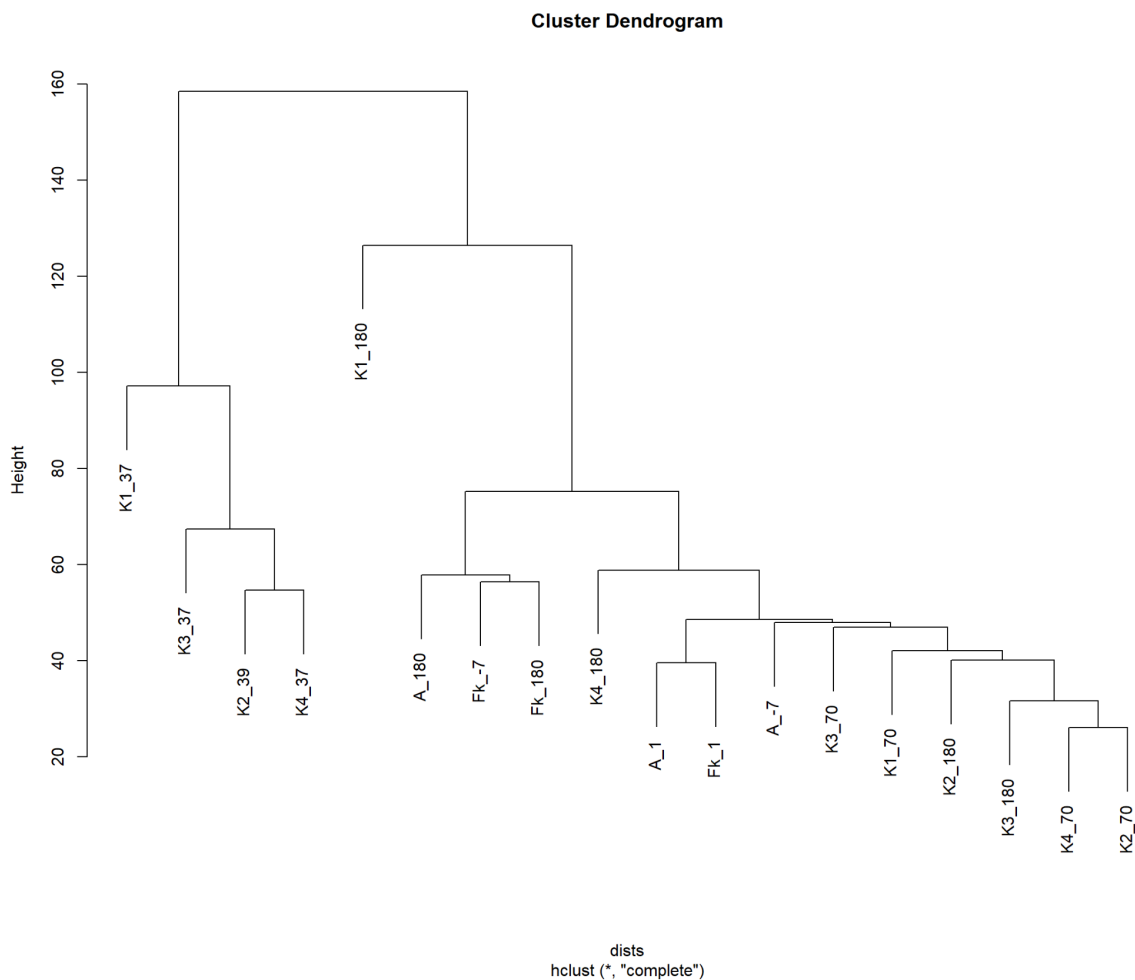
6. ábra: V4 diverzitási indexek

A V4 amplitikon szekvenálással feldolgozott 18 kitüntetett időpontban vett minta (6. ábra) közül a Shannon diverzitásindex szerint 7 mutat 5-nél nagyobb alfa diverzitást, 9 minta diverzitása 4 és 5 közötti értéket vesz fel, 2 minta diverzitásértéke 3,5 és 4 közötti. A Simpson index szerint 12 minta diverzitása 0,98 feletti, illetve 6 minta vesz fel 0,96 és 0,98 közötti értékeket.

Ezen eredmények alapján megállapítható, hogy a V4 amplitikon szekvenálási módszerrel vizsgált minták jelentősen magasabb alfa diverzitási értékeket mutatnak a V1-V3 régiók azonosításával vizsgált mintáknál. Mindazonáltal az alfa diverzitás - mely a minta fajgazdagságának és a fajeloszlás egyenletességének függvénye - értéke függ a vizsgálati módszer által detektált taxonok számától, a módszer érzékenységtől, így az alfa diverzitási értékek önmagukban nem nyújtanak kellően megbízható információt egy adott vizsgálati módszer minőségéről.

3. Klaszter dendrogramok

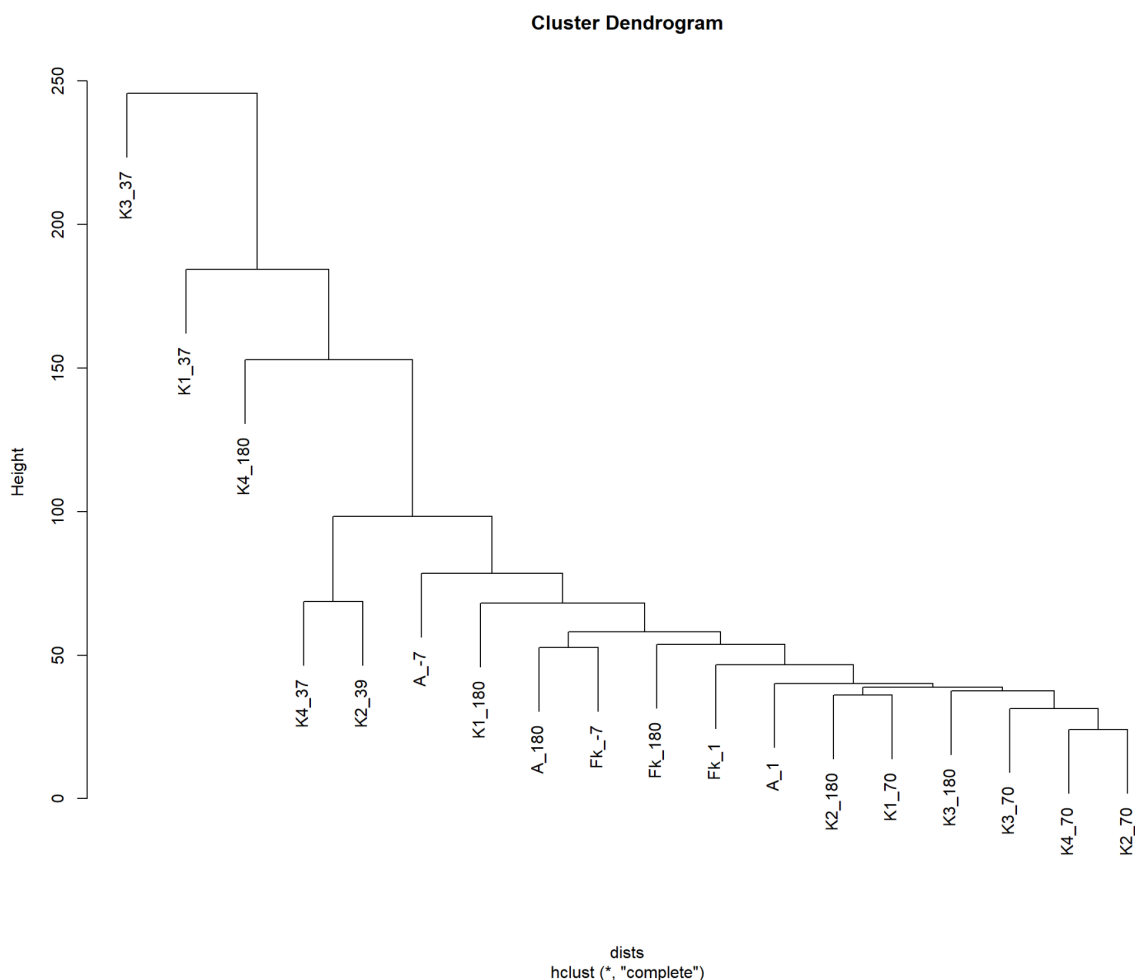
A klaszter dendrogramok az egyes minták egymáshoz való hasonlóságát jelenítik meg vizuálisan. A két szekvenálási megközelítésre alapuló adatok klaszter dendrogramja eltérő képet mutat a minták hasonlósági viszonyairól.



7. ábra: V4 dendrogram

A V4 szekvenálás dendrogramján (7. ábra) az öthetes kölyök minták egy, a többi mintától nagymértékben eltérő klasztert alkotnak. A tíz hetes mintákat tartalmazó klaszterben 2 kölyök 180 napos mintája is megjelenik - ez arra utalhat, hogy az állatok mikrobiomja a 10. héten már viszonylag stabil, kompozíciója hasonlít a felnőttkori állapotokra. A 10 hetes, illetve a K2_180 és K3_180 mintákkal az anya szülés előtti, illetve a két felnőtt állat szülés utáni mintája mutat nagy hasonlóságot. Az A_1 és Fk_1 minták hasonlósága, valamint az A_-7 mintához való közelségük arra enged következtetni, hogy a szülés nem befolyásolja nagy mértékben az anyai mikrobiomot. A két felnőtt állat hatodik havi (A_180, Fk_180), illetve a kontroll hím -7. napi (Fk_-7) mintái

alkotnak egy újabb szoros klasztert, melyen belül a Fk_-7 és Fk_180 minták mutatják a legnagyobb hasonlóságot. A K4_180 minta a hatodik hónapban már BARF táplálással etetett kölyöktől származik, és bár a többi mintától elkülönülten helyezkedik el a dendrogramon, a legnagyobb hasonlóságot a felnőtt mikrobiomokkal mutatja. A többi mintától a legtávolabb a K1_180 áll, ami meglepő, hiszen az állat a hozzátáplálástól fogva ugyanazt a tápot kapta és a tenyésztői kennelben maradt A és Fk egyedekkel. A táp-és környezeti eredetű hatások A, Fk és K2 egyedek mikrobiomjával mutatott nagyobb hasonlóságot indokolnának.



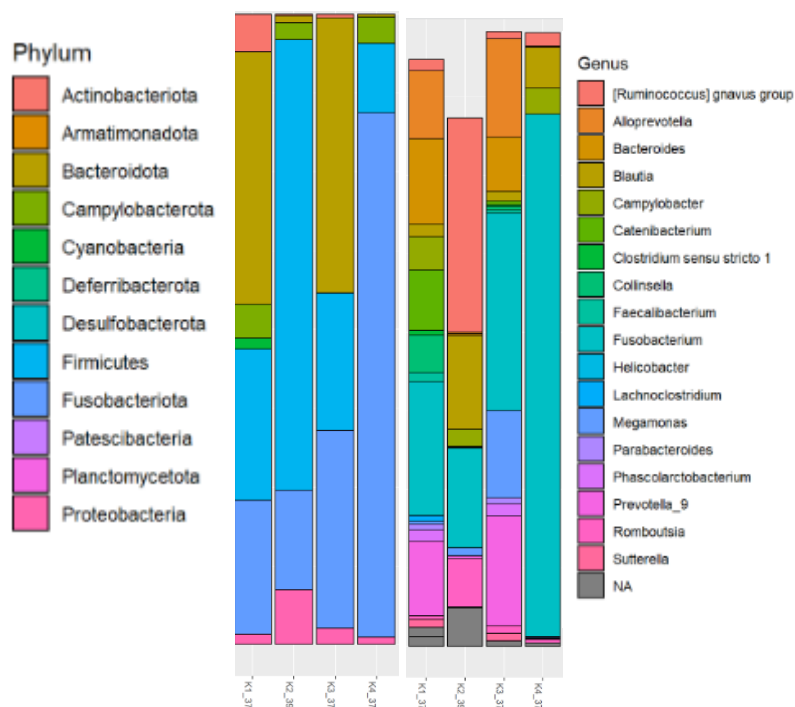
8. ábra: V1-V3 dendrogram

A V1-V3 szekvenálás dendrogramja (8. ábra) szerint az öthetes minták (9-10. ábrák) közül K2_39 és K4_37 nagy hasonlóságot mutat. K1_37 és K3_37 mintákhoz is relatíve közel állnak, ám a V1-V3 szekvenálás nagyobb eltéréseket mutatott ki az öthetes minták között, mint a V4 alapú megközelítés. Az irodalmi adatok alapján a fiatal állatok bélflórája nagymértékben változókéony és

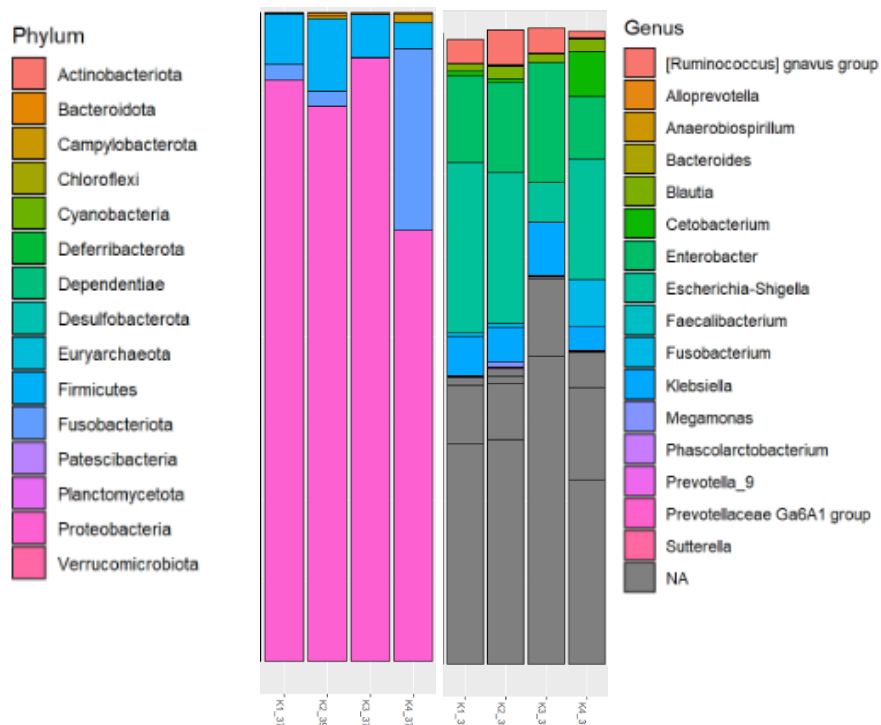
instabil lehet, a közösség diverzitása és kolonizációs ellenállása fiatal korban jellemzően alacsony (Haller, Hornef 2018, Brugiroux és mtsai. 2016, Stetcher és mtsai. 2013). Ezen ismeretek alapján a kölyök mikrobiomok között fiziológiás állapotban inkább az eltérések, mint a nagy hasonlóság feltételezhető.

A dendrogram jobb szélén a négy 70. napi minta a K2_180 és K3_180 mintákkal kiegészülve alkot egy jól elkülöníthető klasztert. E tekintetben a V1-V3 a V4 szekvenálással átfedő eredményeket hozott. A felnőtt minták közül ezen megközelítés szerint is nagy hasonlóságot mutatnak az A_180, Fk_-7 és Fk_180 minták, illetve a Fk_1 és A_1 mikrobiomok. Az A_-7 azaz az anya szülés előtti mintája a többitől ugyanakkor eltér, és a K1_180 mintával közvetlenül nagyobb hasonlóságot mutat, mint a többi felnőtt mikrobiommal. A K1_180 minta jól ágyazódik be a többi közé hasonlóság terén, és a V1-V3 szekvenciák alapján a K4_180 minta mutatja a többi érett mikrobiomtól való legnagyobb eltérést. A fentebb említett táplálásbeli és környezetbeli tényezők ismeretében ez indokolt.

Kutatásunk során mindkét megközelítés alapján elvégeztük a minták részletes longitudinális elemzését törzs és nemzetség taxonómiai szinteken; a V1-V3 amplikon szekvenáláson alapuló elemzést TDK-pályamunka formájában publikáltuk (Henkrich, Járay 2022, SZTE SZAOK). Megfigyeltünk bizonyos markáns különbségeket a két módszerrel végzett longitudinális mikrobiom kompozíciós elemzés eredményei között, melyek között kiemelkedő volt az 5. heti kölyök mikrobiomok összetételére vonatkozó eredmények nagyfokú eltérése. A V1-V3 szekvenálás alapján ezen minták közül kettőt viszonylag egyenletes eloszlásban építettek fel a Bacteroidetes, Firmicutes, Fusobacteria törzsek kisebb mennyiségű Actinobacteriával és Proteobacteriával, a két másik állat közül az egyiknek a Firmicutes, másinak a Fusobacteria dominálta a bélflóráját. Ezzel szemben a V4 amplikon szekvenciák alapján a kölyök mikrobiomok szinte egészét a Proteobacteria törzs tette ki az ötödik héten. A V1-V3 módszerrel szemben V4 szekvenálással emellett sikertelen volt ezen minták több abundáns nemzetségének az azonosítása.



9. ábra: Az 5. heti mintákban V1-V3 szekvenálással detektált törzsek (jobbra) és nemzetségek (balra). Ahol az oszlopdiagramok nem érik el az x tengely maximum értékét, ott a 20 legabundánsabb taxon nem teszi ki a mintában fellelhető taxonok egészét.



10. ábra: Az 5. heti mintákban V4 szekvenálással detektált törzsek (jobbra) és nemzetségek (balra). Ahol az oszlopdiagramok nem érik el az x tengely maximum értékét, ott a 20 legabundánsabb taxon nem teszi ki a mintában fellelhető taxonok egészét.

Az egyes minták összehasonlító elemzése során további jelentősnek mondható eltéréseket találtunk a két módszer adatai közt, azonban ezen elemzések nagy kiterjedésük miatt jelen pályamunka részét nem képezik.

4. Eredményeink összefoglalása

Összegezve, a két 16S rRNS gén amplikon szekvenálási megközelítés közül a V1-V3 módszerrel jóval több ASV-t azonosítottunk, elsősorban az igen kis abundanciával jelenlévő taxonok közül, a V4 szekvenálás viszont több V1-V3 által ki nem mutatott taxont detektált, mint fordítva. Emellett a V4 szekvenciák alapján vizsgált minták jóval nagyobb alfa diverzitást mutattak mind a Shannon, mind a Simpson-index szerint. A V4 szekvenálási adatok klaszter dendrogramja és egyes minták összetétele mutatott bizonyos anomáliákat (K4_180 minta elhelyezkedése a dendrogramon, 5 hetes minták homogén összetétele), melyeket a vizsgálati körülmények nem indokolnak. Ezen esetekről a V1-V3 szekvenálás olyan eredményeket hozott, amely a környezeti változók ismeretében indokoltabbak, reálisabbak, mint a V4-es adatok.

A V1-V3 amplikonokról fontos kiemelni, hogy metodikai limitációk miatt a teljes régió rövid leolvasású szekvenálása nem lehetséges; a génszakaszban található V2 régió egy része a szekvenálás és bioinformatikai szűrés során elveszett. Az azonosítás így tulajdonképpen nem három, hanem kettő és fél variábilis régió szekvenciája alapján történt. A V4 régió teljes egészében megszekvenálható volt.

Mivel a két génszakasz vizsgálata bizonyos mértékben eltérő eredményeket hozott, kutatócsoportunk jelenleg a V1-V9 régiók egyben történő hosszú leolvasású szekvenálására alapuló taxonómiai elemzési módszerekre irányuló kísérleteket végez. Ezen megközelítés mikrobiomikai alkalmazását a referencia szakirodalom is indokolja (Zhang és mtsai. 2018). Emellett az alkalmazott DNS-izolálási protokoll optimalizációjára is végzünk kísérleteket. A négy izolálási és három amplikon szekvenálási módszer természetes, egy állattól származó székletmintákon és ismert összetételű mock mintákon történő alkalmazását és az eredmények részletes kiértékelését követően alkotjuk meg azt a végleges protokollt, melyre a további mikrobiomikai kutatásainkat alapozzuk.

Összefoglalás

Kutatócsoportunk az intesztinális mikrobiom elhízásban és társbetegségeiben betöltött szerepét, illetve a mikrobiom táplálkozásban beállt változásokra adott válaszait vizsgálja.

Munkánk során különböző egészségi állapotú, eltérő diétás-és életmódprogramokat követő humán páciensek mikrobiomját, emellett nagyszámú fajtatiszta, eltérő tápokkal etetett, modellorganizmusként alkalmazott pumi kutya mikrobiomját vizsgáljuk újgenerációs szekvenálási eljárásokkal, székletmintából izolált DNS alapján.

A bélflóra prokarióta tagjainak azonosítására alkalmazott sztenderd módszer a prokarióta 16S rRNS gén variábilis régiók szekvenciájának, mint molekuláris taxonómiai markereknek a használata. Mindazonáltal a 9 variábilis régió kiterjedése, biológiai funkciója és mutációs rátája egymástól eltér, így a különböző V régió szekvenciák különböző taxonösszetételi adatokat mutathatnak egy mintáról.

Jelen munkánk során a 16S rRNS gén V1-től V3-ig tartó szakaszának, illetve a V4 régiónak amplikonjait szekvenáltuk Illumina MiSeq platformon. Az amplikonok izolálása 6 kutya (2 felnőtt és 4 kölyök) 6 hónapon át gyűjtött összesen 20 mintájából történt. V1-V3 szekvenálással 20, V4 szekvenálással 19 mintát elemeztünk. Az amplikon szekvenciákat a DADA2 programmal dolgoztuk fel, mely egyedi amplikon szekvencia variánsként (ASV) kezeli a fellelt szekvenciákat. A szekvenciák taxonómiai azonosítására a SILVA 16S rRNS referencia adatbázist használtuk.

Adataink manuális elemzése során célunk volt megállapítani, hogy a két amplikon szekvenálási módszer által szolgáltatott adatok hogyan és milyen mértékben térnek el egymástól, illetve, hogy a két módszer közül melyik lehet ideálisabb a hosszú távú kutatási alkalmazásra. Eredményeink alapján a V1-V3 szekvenálással sikerült több taxont azonosítani, ám a V4 szekvenálás a vizsgált mikrobiomok nagyobb diverzitását mutatta. Kutatócsoportunk a továbbiakban a V1-V9 régió teljes amplikonjának hosszú leolvasású szekvenálását is teszteli metodikai szempontból.

Mikrobiomikai kutatómunkánk hosszú távú eredményei a táplálkozástudomány és az elhízás biológiájának területén jelenthetnek majd értékes tudományos hozzájárulást, terápiás és prevenciós stratégiák alapját képezhetik, és állategészségügyi relevanciával is bírhatnak.

Irodalomjegyzék

Brugiroux, S., Beutler, M., Pfann, C. *et al.* Genome-guided design of a defined mouse microbiota that confers colonization resistance against *Salmonella enterica* serovar Typhimurium. *Nat Microbiol* 2, 16215 (2017). <https://doi.org/10.1038/nmicrobiol.2016.215>

Bukin, Y., Galachyants, Y., Morozov, I. *et al.* The effect of 16S rRNA region choice on bacterial community metabarcoding results. *Sci Data* 6, 190007 (2019). <https://doi.org/10.1038/sdata.2019.7>

Callahan, B., McMurdie, P. & Holmes, S. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J* 11, 2639–2643 (2017). <https://doi.org/10.1038/ismej.2017.119>

Chakravorty, S., Helb, D., Burday, M., Connell, N., & Alland, D. (2007). A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *Journal of microbiological methods*, 69(2), 330–339. <https://doi.org/10.1016/j.mimet.2007.02.005>

Coelho, L.P., Kultima, J.R., Costea, P.I. *et al.* Similarity of the dog and human gut microbiomes in gene content and response to diet. *Microbiome* 6, 72 (2018). <https://doi.org/10.1186/s40168-018-0450-3>

DeGruttola, A. K., Low, D., Mizoguchi, A., & Mizoguchi, E. (2016). Current Understanding of Dysbiosis in Disease in Human and Animal Models. *Inflammatory bowel diseases*, 22(5), 1137–1150. <https://doi.org/10.1097/MIB.0000000000000750>

Haller, D. és munkatársai (2018) The Gut Microbiome in Health and Disease. Springer.
Hivatkozott fejezetek:

Microbiome and Early Life. Hornef, M.

Henkrich, M., Járay, T. A bél mikrobiom öröklődésének és korai változásainak vizsgálata fajtatiszta kutya modellen. SZTE SZAOK TDK pályamunka, 2022. Témavezetők: dr. habil. Tombácz Dóra egyetemi docens és Gulyás Gábor PhD-hallgató, SZTE SZAOK Orvosi Biológiai Intézet.

Iacobini C, Pugliese G, Blasetti Fantauzzi C, Federici M, Menini S. Metabolically healthy versus metabolically unhealthy obesity. *Metabolism*. 2019 Mar;92:51-60. doi: 10.1016/j.metabol.2018.11.009. Epub 2018 Nov 17. PMID: 30458177.

Kerrigan, Z, Kirkpatrick, J. B., D'Hondt, S. Influence of 16S rRNA Hypervariable Region on Estimates of Bacterial Diversity and Community Composition in Seawater and Marine Sediment. *Frontiers in Microbiology*, Volume 10, 2019. DOI=10.3389/fmicb.2019.01640

Li, J., Jia, H., Cai, X. *et al.* An integrated catalog of reference genes in the human gut microbiome. *Nat Biotechnol* 32, 834–841 (2014). <https://doi.org/10.1038/nbt.2942>

Liébana-García, R., Olivares, M., Bullich-Villarubias, C., López-Almela, I., Romani-Pérez, M., Sanz, Y. (2021) The gut microbiota as a versatile immunomodulator in obesity and associated metabolic disorders. *Best Practice & Research Clinical Endocrinology & Metabolism*. Volume 35, Issue 3, 101542, ISSN 1521-690X, <https://doi.org/10.1016/j.beem.2021.101542>

McMurdie PJ, Holmes S (2013) phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLoS ONE 8(4): e61217. <https://doi.org/10.1371/journal.pone.0061217>

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013 Jan;41(Database issue):D590-6. doi: 10.1093/nar/gks1219. Epub 2012 Nov 28. PMID: 23193283; PMCID: PMC3531112.

Ravi, R.K., Walton, K., Khosroheidari, M. (2018). MiSeq: A Next Generation Sequencing Platform for Genomic Analysis. In: DiStefano, J. (eds) Disease Gene Identification. Methods in Molecular Biology, vol 1706. Humana Press, New York, NY. https://doi.org/10.1007/978-1-4939-7471-9_12

Schluenzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., Bashan, A., Bartels, H., Agmon, I., Franceschi, F., Yonath, A. Structure of Functionally Activated Small Ribosomal Subunit at 3.3 Å Resolution, Cell, Volume 102, Issue 5, 2000, Pages 615-623, ISSN 0092-8674, [https://doi.org/10.1016/S0092-8674\(00\)00084-2](https://doi.org/10.1016/S0092-8674(00)00084-2).

Schuwirth, B. S., Borovinskaya, M. A., Hau, C. W., Zhang, W., Vila-Sanjurjo, A., Holton, J. M., Doudna Cate, J. H. Structures of the Bacterial Ribosome at 3.5 Å Resolution. Science. Volume 310, number 5749, pages 827-834, year 2005. doi = {10.1126/science.1117230}

Sender, R., Fuchs, S., Milo, R. (2016) Revised estimates for the number of human and bacteria cells in the body. PLOS Biology. <https://doi.org/10.1371/journal.pbio.1002533>

Stetcher, B., Berry, D., Loy, A. (2013) Colonization resistance and microbial ecophysiology: using gnotobiotic mouse models and single-cell technology to explore the intestinal jungle. FEMS Microbiology Reviews, Volume 37, Issue 5, September 2013, Pages 793–829, <https://doi.org/10.1111/1574-6976.12024>

Thursby, E., & Juge, N. (2017). Introduction to the human gut microbiota. *The Biochemical journal*, 474(11), 1823–1836. <https://doi.org/10.1042/BCJ20160510>

Turnbaugh, P., Ley, R., Hamady, M. *et al.* The Human Microbiome Project. *Nature* 449, 804–810 (2007). <https://doi.org/10.1038/nature06244>

Vilela de Oliveira, G. L. Chapter 33 - The Gut Microbiome in Autoimmune Diseases, Editor(s): Joel Faintuch, Salomao Faintuch, Microbiome and Metabolome in Diagnosis, Therapy, and other Strategic Applications, Academic Press, 2019, Pages 325-332, ISBN 9780128152492, <https://doi.org/10.1016/B978-0-12-815249-2.00033-6>.

Wilkins, L.J., Monga, M. & Miller, A.W. Defining Dysbiosis for a Cluster of Chronic Diseases. *Sci Rep* 9, 12918 (2019). <https://doi.org/10.1038/s41598-019-49452-y>

Zhang J, Ding X, Guan R, Zhu C, Xu C, Zhu B, Zhang H, Xiong Z, Xue Y, Tu J, Lu Z. Evaluation of different 16S rRNA gene V regions for exploring bacterial diversity in a eutrophic freshwater lake. *Sci Total Environ.* 2018 Mar 15;618:1254-1267. doi: 10.1016/j.scitotenv.2017.09.228. Epub 2017 Oct 28. PMID: 29089134.

Ábrák forrásai

1. ábra:
<https://biology.stackexchange.com/questions/54823/what-causes-the-variable-conserved-structure-in-the-16s-rna-gene>
2. ábra: <https://blogs.iu.edu/ncgas/2021/09/17/introduction-to-illumina-sequencing/>
3. ábra:
<https://www.illumina.com/science/technology/next-generation-sequencing/plan-experiments/paired-end-vs-single-read.html>
- 4- 10. ábrák: saját ábrák (MultiQC, Phyloseq)

Függelék

1. függelék: a vizsgált állatok által fogyasztott kutyatápok analitikai összetétele
 - a. Felnőtt táp (A és Fk által fogyasztva): Brit Care Dog Show Champion Salmon & Rice
 - i. Összetevők: Szárított lazac (35%), rizs (32%), szárított hering (10%), csirkezsír (tokoferollokkal tartósítva), lazacfehérje (5%), lazacolaj (4%), szárított alma, természetes aromák, sörélesztő, ligetszépe olaj (1%), meszes héjú állatok hidrolizált héja (glükózamin-forrás, 320 mg/kg), porckivonat (kondroitin-forrás, 190 mg/kg), mannán-oligo-szacharidok (180 mg/kg), gyógynövények és gyümölcsök (rozmaring, szegfűszeg, citrus, kurkuma, 180 mg/kg), frukto-oligo-szacharidok (120 mg/kg), yucca schidigera (120 mg/kg), inulin (110 mg/kg), máriatövis (90 mg/kg).
 - b. Kölyök táp 1 (a hozzátáplálást követően az összes kölyök által, K1-K2 esetén a vizsgálat teljes időtartama alatt fogyasztva): Brit Care Grain Free Puppy Salmon & Potato
 - i. Összetevők: Szárított lazac (35%), burgonyafélék (28%), lazacprotein (15%), csirkezsír (tokoferollokkal tartósítva), szárított alma, természetes ízesítők, lazacolaj (3%), sörélesztő, hidrolizált rákhéj (glükózamin-szulfát forrás, 320 mg/kg), porckivonat (kondroitin-szulfát forrás, 190 mg/kg), mannán-oligoszacharidok (180 mg/kg), fűszernövények és gyümölcsök (rozmaring, szegfűszeg, citrusfélék, kurkuma, 180 mg/kg), frukto-oligoszacharidok (120 mg/kg), yucca schidigera kivonat (120 mg/kg), inulin (110 mg/kg), máriatövis (90 mg/kg).
 - c. Kölyök táp 2 (K3 által 2 hónapos kortól fogyasztva) Brit Care Puppy All Breed Lamb & Rice
 - i. Összetevők: Bárányhúsliszt (45%), rizs (30%), csirkezsír (tokoferollokkal tartósítva), szárított alma, lazacolaj (3%), természetes aromák, sörélesztő, meszes héjú állatok hidrolizált héja (glükózamin-forrás, 320 mg/kg), porckivonat (kondroitin-forrás, 190 mg/kg), mannán-oligo-szacharidok (180 mg/kg), gyógynövények és gyümölcsök (rozmaring, szegfűszeg, citrus, kurkuma, 180 mg/kg), frukto-oligo-szacharidok (120 mg/kg), yucca schidigera (120 mg/kg), inulin (110 mg/kg), máriatövis (90 mg/kg).
 - d. BARF étrend (K4 által 2,5 hónapos kortól fogyasztva)

i. 75% nyers hús (pulyka, csirke, nyúl, bárány, ló, belsőségek), 25% nyers zöldség és gyümölcs (cukkini, tök, uborka, répa, brokkoli, áfonya, alma)

2. függelék: a V1-V4 amplikon szekvenálással is detektált taxonok: az *Actinobacteria*, *Bacteroidetes*, *Campylobacterota*, *Cyanobacteria*, *Deferribacterota*, *Desulfobacterota*, *Firmicutes*, *Fusobacteria*, *Patesibacteria*, *Planctomycetota*, *Proteobacteria* törzsek, az *Actinobacteria*, *Alphaproteobacteria*, *Bacillii*, *Bacteroidia*, *Campylobacteria*, *Clostridia*, *Coriobacteria*, *Cyanobacteria*, *Deferribacteres*, *Desulfovibrionia*, *Fusobacteria*, *Gammaproteobacteria*, *Negativicutes*, *Vampirivibrionia* osztályok, az *Acholeplasmatales*, *Acidaminococcales*, *Bacteroidales*, *Burkholderiales*, *Campylobacteriales*, *Clostridia* UCG-014, *Clostridiales*, *Coriobacteriales*, *Enterobacteriales*, *Erysipelotrichales*, *Fusobacteriales*, *Gastranaerophilales*, *Lachnospirales*, *Lactobacillales*, *Oscillospirales*, *Peptococcales*, *Peptostreptococcales*-*Tissierelales*, *Propionibacteriales*, *Rhodospirillales*, *Veilonellales*-*Selenomonadales* rendek, az *Acidaminococcaceae*, *Bacteroidaceae*, *Campylobacteraceae*, *Coriobacteraceae*, *Enterobacteriaceae*, *Erysipelotrichaceae*, *Fusobacteriaceae*, *Helicobacteraceae*, *Lachnospiraceae*, *Muribaculaceae*, *Oscillospiraceae*, *Peptostreptococcaceae*, *Prevotellaceae*, *Ruminococcaceae*, *Selenomonadaceae*, *Sutterellaceae*, *Tannerellaceae* családok és a *Ruminococcus gnavus* group, *Alloprevotella*, *Bacteroidetes*, *Blautia*, *Faecalibacterium*, *Fusobacterium*, *Megamonas*, *Phascolarctobacterium*, *Prevotella*, *Sutterella* nemzetségek.

3. függelék: a kizárólag V1-V3 amplikon szekvenálással detektált taxonok: az *Amatimonadota* törzs, az *Amatimonadia*, *Parcubacteria*, *Phycisphaerae* és további beazonosítatlan osztályok, a *Peptococcales* és *Propionibacteriales* rendek, a *Clostridiaceae*, *Erysipelatoclostridiaceae*, *Lactobacillaceae* családok, a *Campylobacter*, *Catenibacterium*, *Clostridium sensu stricto*, *Colinsella*, *Lachnoclostridium*, *Ligilactobacillus*, *Parabacteroides*, *Rombutsia* és további beazonosítatlan nemzetségek.

4. függelék: a kizárólag V4 amplikon szekvenálással detektált taxonok: a *Chloroflexi*, *Dependentiae*, *Desulfobacterota*, *Euryarchaeota*, *Verrucomicrobiota* törzsek, az *Anaerolineae*, *Baeliae*, *Methanobacteria*, *Planctomycetes*, *Saccharimonadia*, *Verrucomicrobiae* osztályok, a *Corneybacteriales*, *Desulfovibrionales*, rendek, a *Succinivibrionaceae* és további beazonosítatlan családok, az *Anaerobispirillum*, *Cetobacterium*, *Enterobacter*, *Escherichia-Shigella*, *Kleibsell*, *Prevotellaceae* *Ga6A1* group és további beazonosítatlan nemzetségek.

Köszönetnyilvánítás

A pályamunka végére érve szeretném megköszönni Járay Tamásnak a wet lab munkafolyamatok elvégzését, Gulyás Gábornak és Kakuk Balázsnak a szekvenciaadatok bioinformatikai elemzését, dr. Tombácz Dórának a kutatás koordinálását és prof. Boldogkői Zsoltnak az intézeti háttér biztosítását. Szeretném megköszönni a Duna-menti Dumás Kennel tenyésztőinek és a kiskutyák új gazdáinak a kutatásban való részvételt és a mintagyűjtésben történő együttműködést.

Témavezetőimnek és a projekten dolgozó munkatársaknak nagyon hálás vagyok minden segítségért, konzultációért, a munkám figyelemmel kíséréséért és a részvételi lehetőségért.

A projekt a Magyar Tudományos Akadémia Lendület Program LP2020/8 (Tombácz Dóra) támogatásával jött létre.

TÉTELES SZERZŐI HOZZÁJÁRULÁS*

Alulírott Henkrich Miksa és Járay Tamás nyilatkozunk, hogy az SZTE 2022 tavaszi helyi TDK-ra benyújtott 'Mikrobiom elemzés során alkalmazott 16S rRNS gén V1-V3 és V4 amplikon szekvenálási módszerek összehasonlítása' című dolgozatunkban leírtak esetében az alábbi munkafolyamatok elvégzésében vettünk részt:

Rész/feladat*	Felkészítés után önállóan végeztem a dolgozathoz szükséges mennyiségben (%-ban megadva)	Folyamatos segítséggel végeztem a dolgozathoz szükséges mennyiségben (%-ban megadva)
Kutya székletminták begyűjtése	0	100
A minták előkészítése és feldolgozása, DNS izoláció	100 (J.T.)	0
PCR amplifikációk, könyvtárkészítés, szekvenálás	100 (J. T.)	0

<i>Bioinformatikai elemzések (DADA2, Phyloseq)</i>	<i>0</i>	<i>100</i>
<i>Manuális adatelemzés, adatok kiértékelése</i>	<i>100 (H. M.)</i>	<i>0</i>
<i>Referencia irodalom áttekintése</i>	<i>100 (H. M.)</i>	<i>0</i>
<i>Pályamunka és előadás kidolgozása</i>	<i>100 (H.M.)</i>	<i>0</i>

¹A szerző-hallgató társával együtt végezte a terepi munkát, melynek során a dolgozat alapját képező fajokra vonatkozó anyagok begyűjtését a terület/vizsgálati időszak egyik felében teljes mértékben ő végezte önállóan. A fennmaradó részében az anyaggyűjtést más végezte.

²A felhasznált 5 primer közül a szerző-hallgató 3-at maga tervezett, a többi készen kapta. ³A western-blot reakciók tervezését (pl.: minta sorrend+mennyiség), elvi összeállítását a szerző hallgató maga végezte, ugyanakkor a munka „wet” részét jórészt a laborhoz tartozó asszisztens

végezte

* Az alábbi lista mintául szolgál, szabadon módosítandó/bővíthető

Kelt: Szeged, 2022 április 11

(Kérjük a dőlttel jelölt részek értelemszerű megadását. A táblázat szükség szerint értelemszerűen módosítható).

Többszerzős dolgozat esetén:

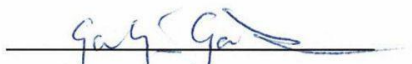


Pályamunkát benyújtó hallgató (Henkrich Miksa) aláírása

A nyilatkozatban foglaltak helyességét megerősítem:



Témavezető (dr. Tombácz Dóra egyetemi docens, SZTE SZAOK Orvosi Biológiai Intézet) aláírása



Témavezető (Gulyás Gábor PhD-hallgató, SZTE SZAOK Orvosi Biológiai Intézet) aláírása



Társszerző (Járny Tamás) aláírása