



НАЦИОНАЛЕН ЦЕНТЪР ПО ЗАРАЗНИ И ПАРАЗИТНИ БОЛЕСТИ

Отдел “Микробиология”,
Лаборатория по “Особено опасни бактериални инфекции” (ООБИ)
Сектор “Микробиом”

**Микробиомен и биоинформатичен анализ за
проучване на патогенезата на саркоидозата**

Йордан Красимиров Ходжев

ДИСЕРТАЦИЯ

за присъждане на образователна и научна степен “Доктор” по
научна специалност 01.06.12 Микробиология

Научен ръководител: проф. дн Стефан Панайотов

София, 2023 г.

Дисертационният труд е изработен в Отдел „Микробиология“ към НЦЗПБ, гр. София с финансовата подкрепа на проект “САРКОИДОЗА”, финансиран от ННП “ВИХРЕН” и договор с ФНИ: КП-06-ДВ/10 от 21.12.2019.

Представен е на разширен колегиум на отдел „Микробиология“ на 25.09.2022 г. и е насочен за защита пред научно жури.

Дисертационният труд съдържа 128 страници, 24 фигури и 10 таблици. Библиографията включва 121 литературни източника.

Материалите във връзка със защитата са публикувани на интернет страницата на НЦЗПБ и се намират на разположение в библиотеката на НЦЗПБ, гр. София

Публичната защита ще се състои на 15.02.2024 г. отч. в аулата на НЦЗПБ, бул. „Янко Сакъзов“ № 26, гр. София, на открито заседание на научното жури, назначено със заповед № 583/01.12.2023г. на Директора на НЦЗПБ, съгласно ППЗРАСРБ на НЦЗПБ и ЗРАСРБ.

1. Въведение

Саркоидозата е рядко възпалително заболяване, с развитие на грануломи в различни органи. Предимно са засегнати белите дробове и лимфните възли. През последните две десетилетия беше постигнат значителен напредък в диагностиката, клиничната оценка и лечението на саркоидозата. За съжаление до този момент точната етиология на заболяването остава неизвестна.

Потенциалното участие на микроби в патогенезата на саркоидозата е предложено преди години. При саркоидоза се наблюдават хистологични прилики с грануломатозните лезии, причинени от патогенни микроорганизми, напр. туберкулоза, бруцелоза и др. Епидемиологични и микробиологични проучвания предполагат, че поне при част от пациентите, микроби или техни продукти могат да предизвикат имунен отговор, водещ до образуване на саркоиден гранулом. Саркоидозата може да наподобява много автоимунни заболявания, и наблюдаваното облекчаване на симптомите в отговор на кортикостероиди или имunosупресивни лекарства, подкрепя идеята, че автоимунната реакция е критичен компонент в патогенезата на това заболяване.

Ролята на микробиома при различни автоимунни и възпалителни заболявания е активна област на изследване (Tsafarova, 2021). Микробиома участва в модулирането на имунната система и повлияване на системното възпаление. Налице са научни доказателства за взаимодействието на микробите с имунните клетки, които могат да участват в имунопатогенезата на саркоидозата. Нови проучванията показват имуномодулиращата роля на белодробния микробиом при саркоидоза. Връзката между микробиома на кръвта и саркоидозата не е проучена. За изясняването на тази многостранна връзка са необходими интердисциплинарни изследвания, комбиниращи микробиологични, имунологични, клинични и биоинформатични знания. Напредъкът в областта на биоинформатиката и изследванията на кръвния микробиом дават нови възможности за изясняване на патогенезата на саркоидозата.

Основна работна хипотеза на настоящето изследване е, че микробният дисбаланс или свързаните количествени изменения на специфични представители на кръвния

микробиом могат да допринесат за развитие на белодробната саркоидоза или дори да я индуцират.

Предложен е иновативен подход за изследване и оценка на участието на кръвния микробиом в имунния и възпалителния отговор, основан върху съвременни и основополагащи научни аргументи, а именно:

(1) Дисбиозата на кръвния микробиом или специфични персистиращи микробни видове могат да предизвикат (хронични) възпалителни реакции, потенциално допринасящи за развитието или обострянето на саркоидозата (Panaiotov et al., 2018). Хроничното възпаление е ключова характеристика на саркоидозата. Предполага се, че дисбалансът на кръвния микробиом би предизвикал необичаен имунен отговор, водещ до болестни или хронични състояния, включително саркоидоза.

(2) Автоимунният отговор към антигеновете на микробни агенти на кръвния микробиом, би допринесъл за образуването на грануломи при саркоидоза.

(3) Метаболитите на кръвния микробиом биха могли да участват в патогенезата или при влошаване развитието на саркоидозата.

(4) Ролята на кръвния микробиом е доказана в патогенезата на множество възпалителни и системни заболявания.

(5) Методите за метагеномен анализ дават възможност за моментен количествен и качествен анализ на микробиомния състав.

(6) Съвременните методи за статистически микробиомен анализ и машинно обучение изследват сложно структурирани данни от единични или малки групи пациенти.

Ролята на кръвния микробиом в патогенезата на множество заболявания е сравнително нова научна направление, което предстои да бъде развита с помощта на новите техники на микробиомен анализ и биоинформатиката.

Настоящите изследвания, обект на дисертационния труд, разкриват детайли относно участието на кръвния микробиом в патогенезата на саркоидозата. Анализът на кръвния микробиом предлага оригинални данни за нови диагностични маркери за идентифициране и наблюдение при саркоидоза.

Получени са оригинални данни, които подпомагат разработването на нови терапевтични стратегии и персонализиране на лечението на грануломатозните белодробни заболявания.

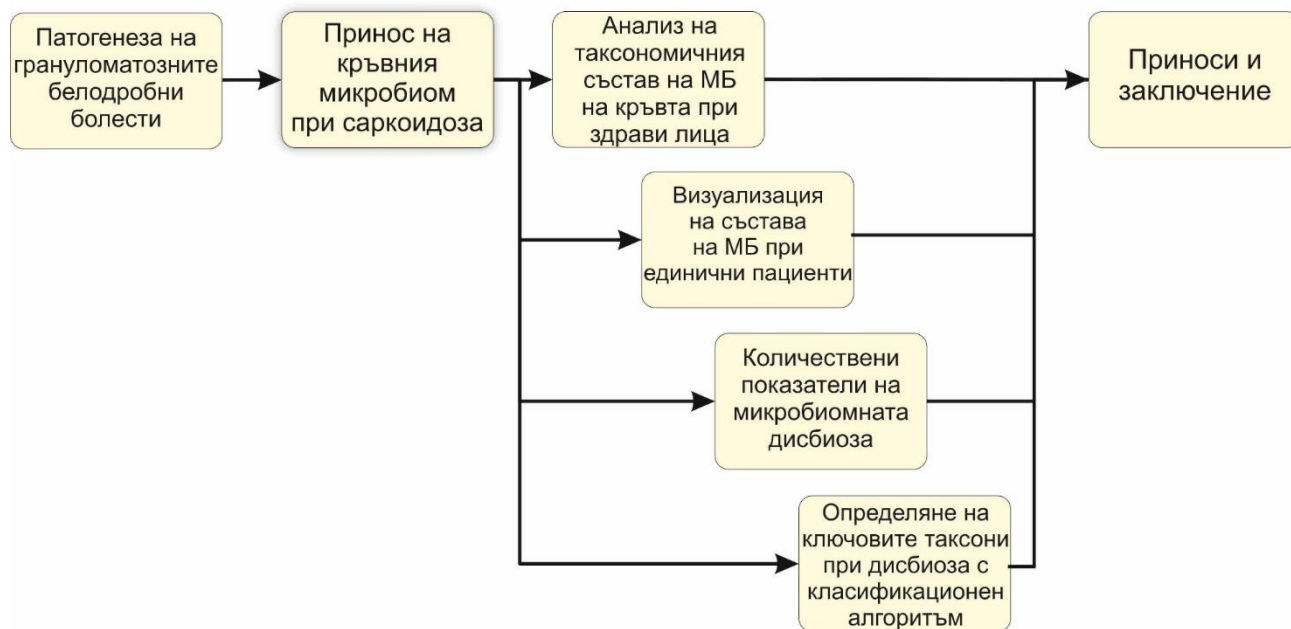
Дисертационният труд включва следната структура (фиг. 1):

(1) **Биоинформатичен анализ** на базисния (core) кръвен микробиом при здрави лица, спрямо който се сравнява кръвния микробиом на пациенти със саркоидоза.

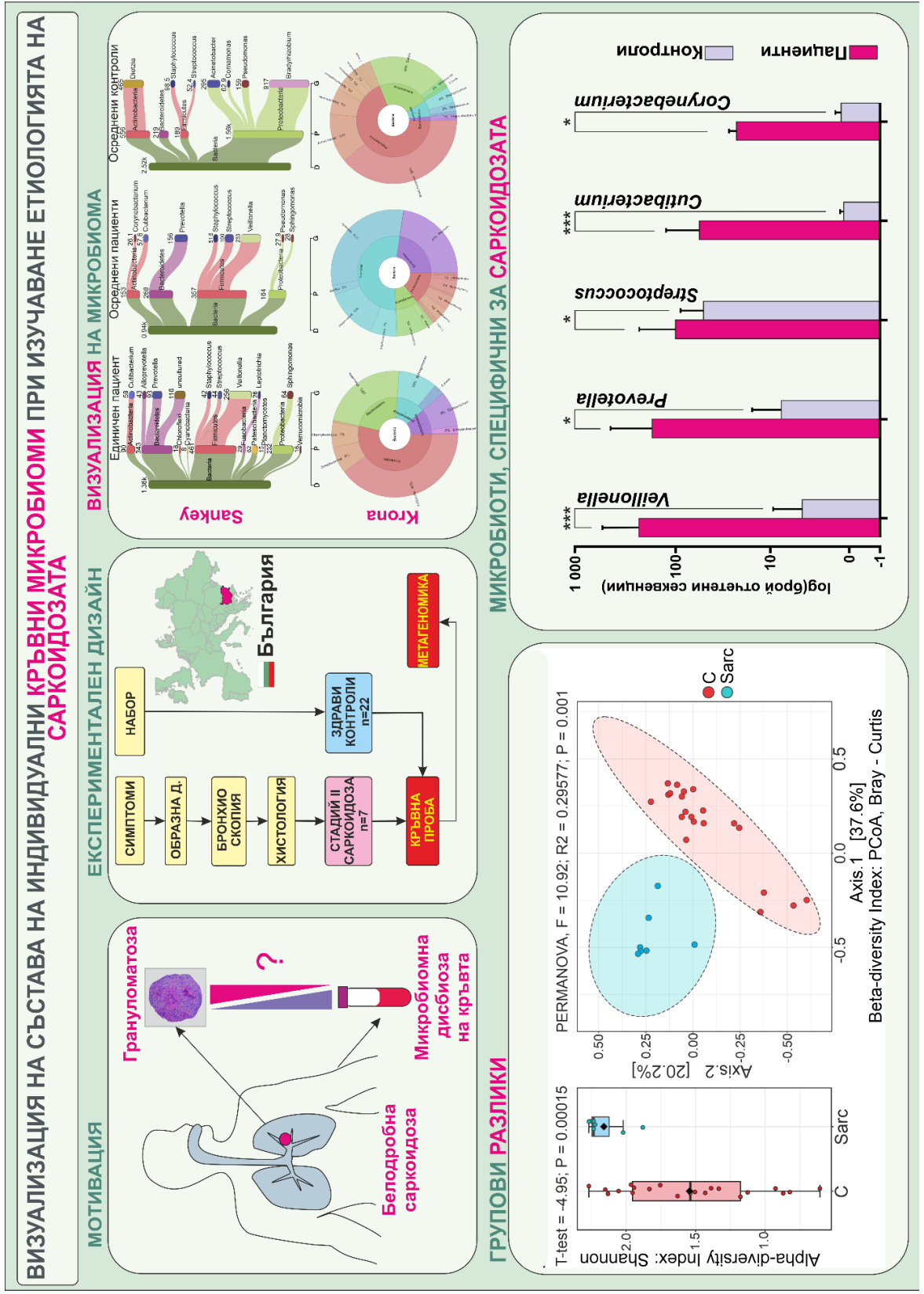
(2) **Визуализация** на кръвния микробиом. Детайлно профилиране на кръвния микробиом при единични пациенти със саркоидоза.

(3) Идентифициране на **количествени показатели на дисбиоза** в кръвния микробиом. Откриване на количествени параметри, които отчитат промени в състава на микробиома, водещи до саркоидозна патология.

(5) **Машинно обучение** – класификация на проби на база на таксономичния състав на кръвния микробиом чрез класификационни дървета. Използване на техники за машинно обучение за идентифициране на вероятни ключови таксони свързани с микробиомна дисбиоза, водеща до развитие на саркоидоза.



Фиг. 1. Структура на дисертационния труд, която следва основните стъпки на биоинформатичния анализ на кръвния микробиом.



Фиг. 2. Схематично представяне на резултатите от дисертационния труд за връзката между състава на кръвния микробиом

2. Цел и задачи на дисертационния труд

2.1. Цел

Целта на настоящия дисертационен труд е провеждане на микробиомен и биоинформатичен анализ за установяване на ролята на кръвния микробиом в патогенезата на белодробната саркоидоза.

2.2. Задачи

За постигане на целта бяха формулирани следните задачи за изпълнение:

2.1.1. Апробиране на биоинформатичен подход за установяване на състава на кръвния микробиом в норма при здрави лица.

2.1.2. Визуализация и оценка на количествения и качествен състав на кръвния микробиом на единични пациенти със саркоидоза.

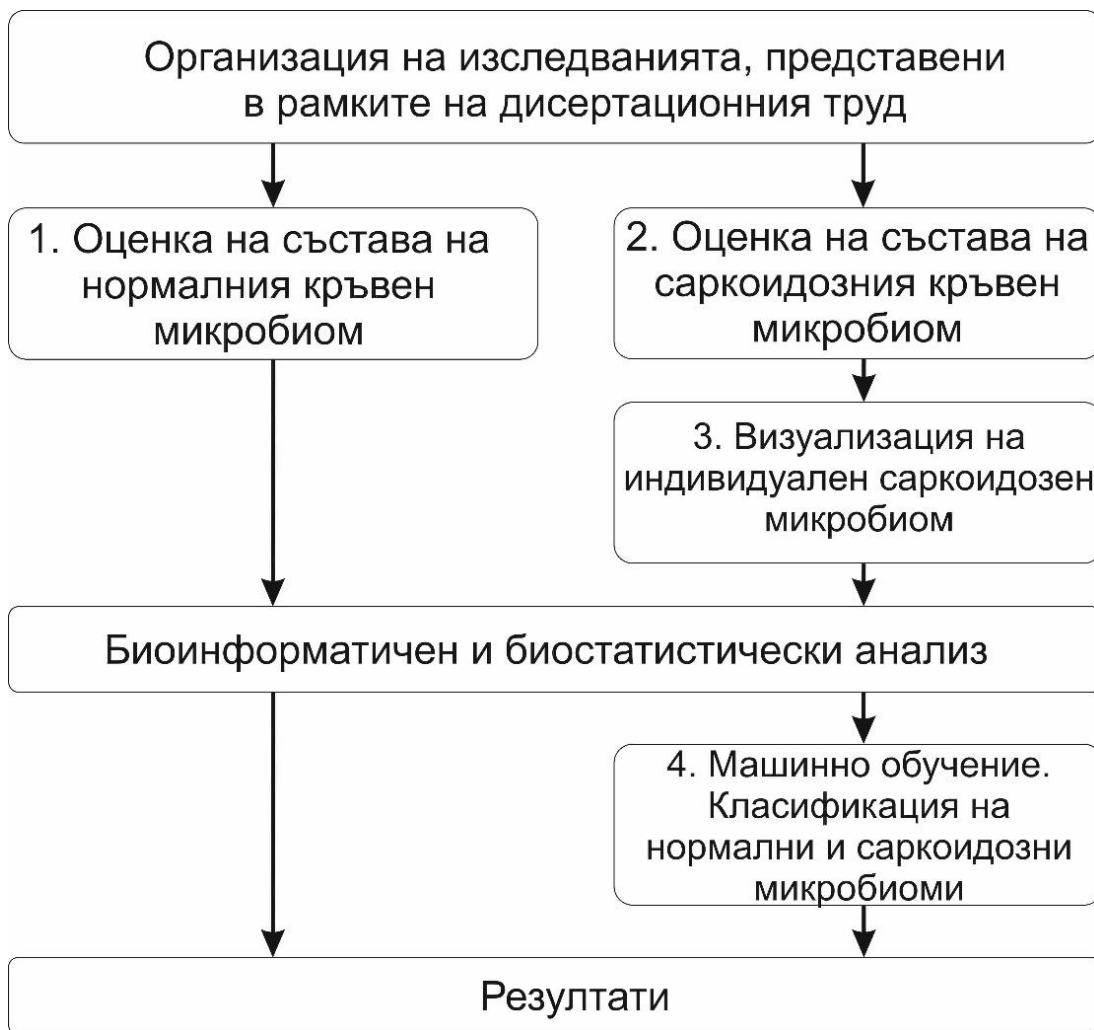
2.1.3. Изясняване на потенциалните причинно-следствени връзки между дисбиоза на микробиома на кръвта и патогенезата на белодробната саркоидозата.

2.1.4. Прилагане на алгоритми за машинно обучение за определяне на ключови микробни таксони с потенциален принос в патологията на белодробната саркоидоза.

3. Материали и методи

3.1. Структура на изследването

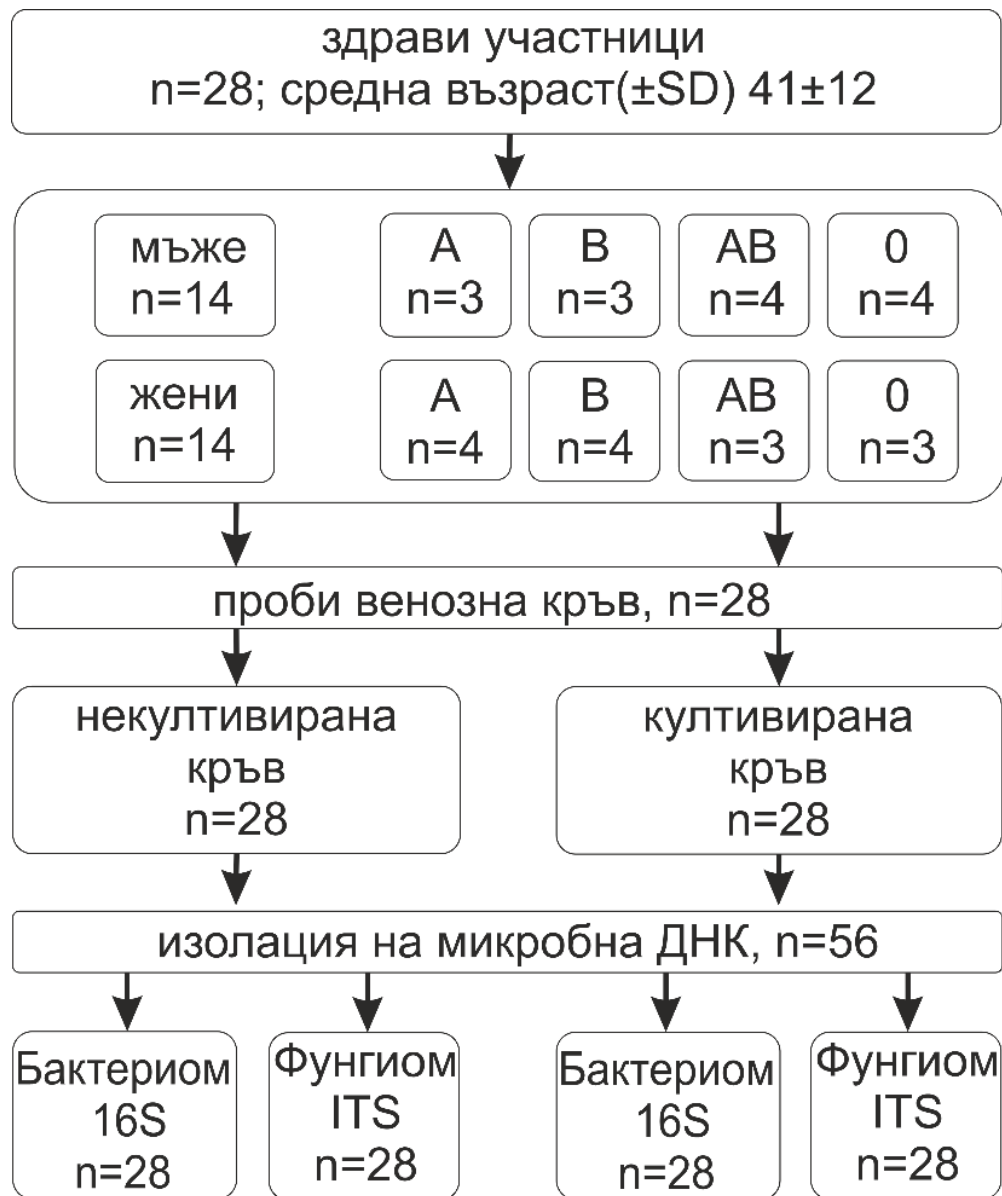
Фигура 3. представя структура и последователност на изследванията, включени в настоящия дисертационен труд.



Фиг. 3. Диаграма на проведените изследвания

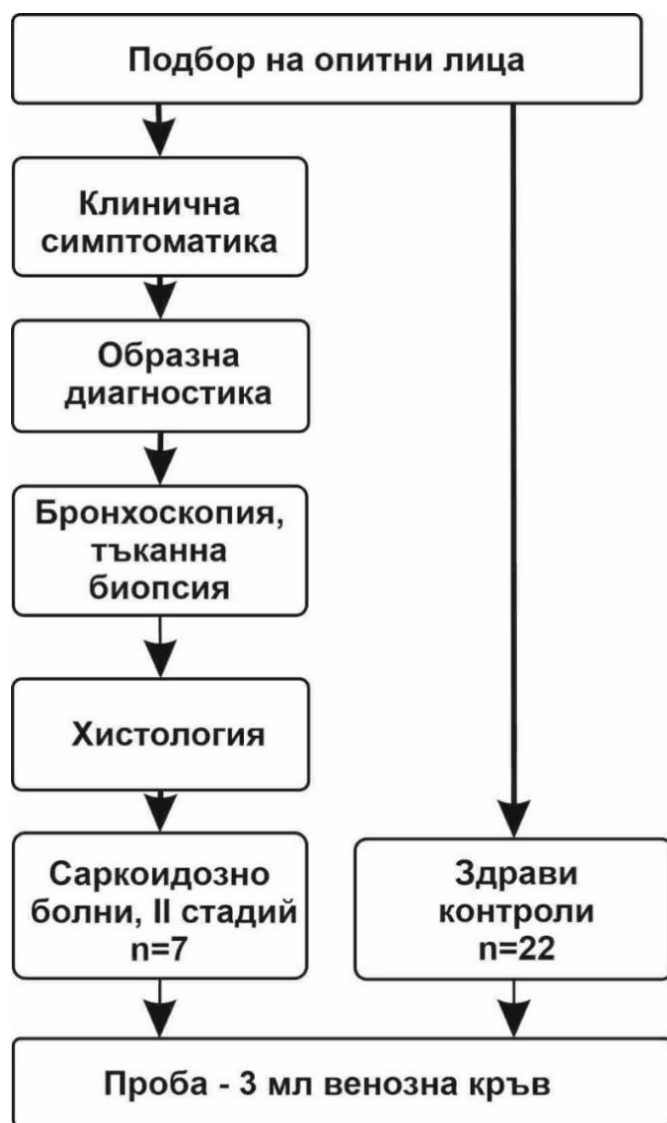
3.1.1. Подбор на контролни лица

При изследване на таксономичния състав на кръвния микробиом бяха подбрани 28 здрави възрастни лица (Фиг. 4). Извадката е балансирана по пол, възраст и кръвна група (Panaiotov et al., 2021).



Фиг. 4. Изследване на кръвния микробиом при здрави лица чрез култивирана и некултивирана венозна кръв.

3.1.2. Подбор на пациенти със саркоидоза за изследване.



Фиг. 5. Схема на изследването.

Между ноември 2020 г. и ноември 2022 г. в отделението по бронхология на УМБАЛ „Св. Иван Рилски“ бяха изследвани повече от 45 пациенти с белодробни оплаквания, акто 7 от тях бяха диагностицирани с белодробна саркоидоза (Hodzhev et al., 2023). За доказването им бяха приложени диагностичните критерии за саркоидоза (Фиг. 5). Пациентите бяха подложени на трансbronхиална биопсия, последвана от хистопатологична

оценка за доказване на не-казеиращи грануломи. Всички изследвани пациенти (Таблица 1) бяха с положителна диагноза за белодробна саркоидоза втори стадий.

Таблица 1. Характеристики на изследваните лица.

Характеристики	Саркоидоза	Клинично здрави лица
Количество проби	7	22
Средна възраст (\pm SD)	48,33 (\pm 19,14)	49,21 (\pm 17,45)
Мъже/жени	4/3	12/10
Разпределение по кръвна група A/B/AB/0	2/2/1/2	7/4/6/5
Саркоидоза тип II	7	0
Брой хистологично потвърдени проби	7	-
Кортикостероидна терапия	0	0
Никога/бивши/настоящи пушачи	4/1/2	10/4/2
Придружаващи заболявания – АХ/СН/ЗД/НП	4 4	1 0 0 0 0

Съкращения: АН - Артериална хипертония; СН - Струма нодоза; СА - Сънна апнея; ЗД - Захарен диабет; НП - неприложимо.

3.2. Събиране и обработка на кръвните проби

Подбраните клинично здравите лица отговаряха на следните условия:

- не е било хоспитализирано през последните две години,
- не е приемало антибиотици и
- не е било подлагано на хирургична стоматологична интервенция през последните шест месеца.

Кръвта беше взета в епруветки Vacutainer сК3ЕDТА (BD, Franklin Lakes, NJ, USA). Кръвните проби бяха разделени на две части. Една част за култивиране (3 ml) и друга част за директно изолиране на ДНК (3 ml). Всички кръвни проби бяха изследвани за стерилитет чрез култивиране на Sabouraud и кръвен агар (BB-NCIPD, София).

3.2.1 Култивиране на кръвните проби

Беше приложен модифициран метод за реанимиране на кръвните микробиоти, разработен от доц. Емил Калфин (Kalfin, 1997). Кръвна проба (0,3 ml) беше добавена към

2,7 ml течна среда ВНІ. Базовата среда се състои от течна среда Brain Heart Infusion (ВНІ), (BD, Франклин Лейкс, Ню Джърси, САЩ), 0,2% дрождев екстракт (BD, Франклин Лейкс, Ню Джърси, САЩ). Базовата среда беше автоклавирана за 20 мин. при 1.5 atm. Към средата беше добавена стерилна (D+) захароза при 10% крайна концентрация и синтетичен, водоразтворим витамин К3 (менадион натриев бисулфит, SigmaAldrich, Бърлингтон, Вермонт, САЩ) в концентрация от 1 mg/ml. Витамин К3 беше стерилизиран чрез ултрафилтрация през 0.22 μm филтър и добавен към основната среда. Култивирането беше извършено в стерилни полипропиленови епруветки от 10 ml Falcon (Corning Inc, Corning, Ню Йорк, САЩ). Култивирането беше извършено при 43°C за 24 часа. Изолираната кръвна микрофлора беше потвърдена чрез оцветяване по Грам.

3.3. Изолиране и секвениране на ДНК

3.3.1. Изолиране и секвениране на ДНК от култивирани и некултивирани кръвени проби.

Пробите от некултивирана цяла кръв бяха лизирани с добавяне на 5 обема стерилна dH₂O. Инкубирането беше за един час на стайна температура. Пробите бяха центрофугирани на 3000 об/мин за 15 мин. Утайката беше трикратно инкубирана и промита за 30 минути при 37°C с 10 ml DNase I 1 U/ml (Thermo Fisher Scientific, Waltham, MA, USA) в ddH₂O. Инкубирането беше с цел да се елиминират остатъчните замърсявания от човешка ДНК, след което се пристъпи към изолиране на ДНК от утайката.

Епруветките с култивираните проби бяха центрофугирани при 3000 g в продължение на 15 минути. Супернатантата беше отлята. Тъмнокафявата утайка беше инкубирана и промита с 10 ml DNase I 1 U/ml (Thermo Fisher Scientific, Waltham, MA, USA) в ddH₂O за 30 минути при 37°C, за да се елиминират остатъчните ДНК замърсявания свързани с лизирани ядрени кръвни клетки, ДНК от хранителната среда или ДНК следи във водата. Клетъчната утайка беше ресуспендирана в 1 ml лизис буфер (500 mM NaCl, 50 mM Трис - HCl, pH 8,0, 50 mM EDTA и 4% натриев додецил сулфат). Същата процедура беше приложена за изолиране на ДНК на микрофлора от цяла кръв, биопсия, БАЛ, смив и котрола от вода и реактиви.

Всеки от подготвените клинични материали беше хомогенизиран в епруветка с циркониеви перли с големина 0.1/0.3 μm (1:1) и лизиращ буфер за 3 мин, на максимална скорост с апарат BeadBug Microtube homogenizer. След това пробите бяха инкубирани за 15 мин при 70°C с неколkokратно разклащане на ръка. Следваше центрофугиране при 4°C за 10 мин при 12 000 об/мин и прехвърляне на супернатантата в нова 2 ml епруветка с перли. Към лизата бяха добавени 300 μl свеж лизиращ буфер с посочения по-горе състав, но без протеиназа К. Стъпките на хомогенизиране, инкубиране и центрофугиране бяха повторени. Следваше прехвърляне на супернатантата в нова 2 ml центрофужна епруветка (без перли), към която бяха добавени 260 μl 10 M амониев ацетат. Пробите бяха вортексирани и инкубирани на 4°C за 15-20 мин за преципитиране на клетъчните отломки, протеини и остатъчният SDS. След центрофугиране на 4°C за 15 мин при 12 000 об/мин супернатантата беше разделена поравно в две 1.5 ml центрофужни епруветки към които беше добавен равен обем изопропанол. Епруветките бяха вортексирани и оставени на -20°C за цяла нощ за утаяване на ДНК. На сутринта пробите бяха центрофугирани при 4°C за 20 мин при 12 000 об/мин, след което супернатантата бе отстранена внимателно, а утаените нуклеинови киселини бяха промити с 500 μl 70% етанол и отново центрофугирани при 4°C за 15 мин при 12 000 об/мин. Супернатантата бе изсмукана внимателно, а утайката изсушена на стайна температура. Утайката от нуклеинови киселини беше разтворена в 40-50 μl стерилна дестилирана вода и разделените проби бяха обединени в обща епруветка. Количеството и чистотата на така изолираната ДНК беше измерена спектрофотометрично с апарат GeneQuant или BioDrop и пробите бяха замразени и съхранявани при -20°C.

По описания метод ДНК беше изолирана от всички клинични метериали (цяла кръв, биопсия, БАЛ, смив и котрола от вода и реактиви). В хода на работа протоколът претърпя някои изменения целящи намаляване на количеството на човешката ДНК, РНК и подобряване на чистотата на изолираната ДНК до съотношение A260/A280 над 1.8.

3.3.2. Създаване на таргетни метагеномни библиотеки

За създаване на геномни библиотеки беше използван фирмен кит NEXTFLEX 16S V4 & 18S Amplicon-Sec Kit (PerkinElmer, USA), съдържащ баркодове за 8 проби. Подготвените библиотеки покриват, както бактериалния V4 хипервариабилен участък на

16S rRNA гените, така и 18S регионите, чрез които могат да се идентифицират гъбички и еукариотни микроорганизми.

Геномните библиотеки бяха изготвени следвайки протокола приложен от производителя. Накратко той се базира на две основни PCR реакции и стъпки на почистване на ДНК. При първата бяха амплифицирани едновременно специфичните таргетни региони на 16S V4 и 18S домените. При вторият PCR в краищата на ДНК молекулите бяха добавени фрагменти, които се залавят за поточната клетка (flow cell) при секвенирането и индексите, които са специфични за всяка проба и представляват последователности от 12 нуклеотидни бази. След всяка от амплификациите беше проведена стъпка на почистване на ампликоните от къси ДНК-фрагменти, праймери, индекси и димери, чрез предоставените в кита магнитни частици (Cleanup Beads). Следваше валидиране на библиотеките за всяка от пробите, чрез 2% гел електрофореза. Наблюдаваха се отчетливи ивици с големина 450 нб за 16S V4 региона и 550 нб за 18S. Накрая пробите бяха обединени в общ ДНК пул с концентрация 4,12 nM. Последният беше денатуриран с 0,2 M NaOH, разреден до крайна концентрация от 10 nM и към така приготвеният пул беше добавена вътрешна качествена контрола на секвенирането (5% Phix).

3.3.3. Таргетно метагеномно секвениране

За ДНК секвениране приложихме протоколи за 16S rRNA гени и ITS2 таргетно секвениране на Illumina MiSeq (Illumina Inc., Сан Диего, Калифорния, САЩ). ДНК. Секвенирането беше извършено в IMG/M Laboratories GmbH (IMG/M Laboratories GmbH, Мартинсрайд, Германия) и Novogene UK (Китай). За идентификация на бактериалните V3-V4 хиперпроменливи региони на 16S rRNA гените бяха амплифицирани с универсални китове 314F и 805R даващи ампликони от 465 нуклеотиди. За идентификация на гъбички, са амплифицирани ITS2, ITS3 и ITS4 регионите, като са използвани универсални китове, от които се получават ампликони от 400 bp. Извършено е двупосочно секвениране. И двете срещуположни вериги са с дължина от 300 бази, 2 × 300 bp. Получени бяха около 130 милиона риди с покритие Q30 >80% за проба. Суровите секвенции с лошо качество бяха премахнати от по-нататъчния анализ.

3.4. Биоинформатичен анализ

3.4.1. Биоинформатична подготовка на секвениционните данни

За оценка на ефективността на секвенирането беше използван софтуерът Illumina MiSeq Reporter версия 2.5.1.3 (Illumina Inc., Сан Диего, Калифорния, САЩ) и Illumina Sequence Analysis Viewer версия 2.1.8 (Illumina Inc., Сан Диего, Калифорния, САЩ). Първичният анализ на данните и контрола на качеството, включително обработката на сигнали и де-мултиплексиране бяха извършени с помощта на MiSeq Reporter и CLC Genomics Workbench версия 9.5.3 (QIAGEN, København, Дания). QIAGEN CLC Microbial Genomics Module беше използван за таксономично профилиране на OTU съгласно указанията на производителя (<https://digitalinsights.qiagen.com>). Като първа стъпка се четат правите и обратни последователности от всяка проба, като се използва припокриващата се информация. Двете секвенции бяха обединени. Секвенциите по-малки от 320 базови двойки бяха отстранени, а краищата на по-дългите секвенции бяха изрязани до тази дължина с цел да се гарантира високо ниво на полезна информация при филогенетичната класификация. Химерните последователности, които са резултат от PCR подготовката на библиотеките за секвениране, както и артефактите от самото секвениране, също бяха филтрирани.

3.4.2. Таксономична идентификация на микробния състав.

Суровите данни бяха предварително обработени, за да се премахнат шумът и артефактите. Това включва обработка на суровите данни и йерархична. Получените последователности бяха групирани в Оперативни таксономични единици - OTU. Извършено бе секвенционно клъстериране с индекс на сходство 97% (Blaxter et al., 2005; Sneath & Sokal, 1973). Зададохме праг от ≥ 10 секвенции (reads), за да приемем дадена оперативна таксономична единица за валидена. Всяка група от секвенции с $>97\%$ идентичност беше идентифицирана, като клъстер. От всеки клъстер беше дефинирана една референтна последователност на OTU, която е представена в един ред в пълната таблица с резултати. Таблицата за оперативните таксономични единици (OTU) с информация за разнообразие на микробиома до таксономично ниво род беше генерирана с помощта на пакета Pavian (Breitwieser & Salzberg, 2020). За таксономична идентификация бе използван софтуерът Kraken2 (v2 2.1.2; SILVA database, version 2022-02-02T162959Z), който е достъпен през европейския сървър на платформата за биостатистика Galaxy (The Galaxy Community

et al., 2022) (<https://usegalaxy.eu/>). Освен това, към получените от Kraken 2 таксономични разпределения бе приложена Байесова рекалкулация на многообразието за да се получат коригирани стойности на микробното разнообразие за всяка индивидуална кръвна проба (Lu et al., 2022). Микробните родове, открити в отрицателните контроли (вода, реактиви и въздух), бяха изключени от нататъшен анализ.

3.5. Биостатистика

3.5.1. Биостатистически методи

За сравнение на микробните видовете между култивирани и некултивирани проби приложихме статистически софтуер R, версия 3.6.0. Данните за некултивирани и култивирани групи бактерии и гъбички бяха сравнени с t-тест за независими проби. Прагът на значимостта на статистическите сравнения на средногруповите средни стойности се определи като двустранно ограничена стойност ($p < 0,05$). Беше приложена контрола за фалшиво позитивни тестове при множествени сравнения (FDR) с цел избягване на статистически грешки от I род. Средното съдържание на таксони за всяка проба беше изчислено и изразено, като средна стойност \pm стандартно отклонение (\pm SD, Таблица 4). Анализът и визуализацията на данните за микробиома са извършени посредством интернет базираната платформа „*Microbiome Analyst*“ (<https://www.microbiomeanalyst.ca>). Платформата дава възможност за профилиране на ампликонни секвенционни данни, като предоставя онлайн приложение за графичен потребителски интерфейс, R версия 3.5.1 (<https://github.com/xia-lab/MicrobiomeAnalystR>) и осигурява многобройни алгоритми за анализ на метагеномна информация. Таблиците с количествените данни за OTU и съответните метаданни бяха качени на сървър за по-нататъшна обработка. Разнообразието на OTU е определено като брой секвенционни прочита без по-нататъшна трансформация и нормализация. Йерархичното групиране използва два алгоритъма за калкулация на статистическите дистанции по методите на Евклид и Ward. За статистическа оценка на бета анализа е използвана пермутация ANOVA (с пермутация (PERMANOVA)). Сравнението между пробите е извършено чрез клъстерен анализ на базата на анализ на главните координати (PCoA), както и с използване на метода за дивергенция Jensen - Shannon за определяне на разстояние и пермутация ANOVA (PERMANOVA) за статистическо валидиране. Бяха генерирани двумерни PCoA графики за сравнение на култивирани и

некултивирани проби чрез изчисляване на първите два основни компонента, които след това служат, като основни координати на тези графики. Беше извършена топлинна визуализация посредством (Heat map) и PCoA със софтуерната библиотека Microbiome Analyst.

3.5.5. Машинно обучение

Приложихме контролиран анализ на машинното обучение с цел да се обогати информацията за микробиомните промени при саркоидозно болни, получена чрез простото групово сравнение (Hodzhev, 2023). За тази цел приложихме модел на дървото на решенията за класификация. Моделът класифицира пробите на база на техните OTU данни на ниво тип и род в два класа, т.е. саркоидозна спрямо контролна група. Алгоритмите за класификационно дърво често се използват за анализ на сложни, ненормално разпределени данни. Избрахме тази методология поради нейния капацитет за управление на по-малки размери на извадката и нейния агностицизъм при разпространението на данни. Машинното обучение беше приложено в Python версия 3.11.4, като се използва библиотеката scikit-learn. За двоична класификация примесът Gini е мярка за вероятността случайно избран случай да бъде неправилно класифициран.

4. Резултати и обсъждане

С цел следване на логическа последователност на описанието, след всяка изпълнена задача е направено обсъждане на резултатите.

4.1. Многообразие при нормален кръвен микробиом на клинично здрави лица

Чрез култивиране на кръвни проби от здрави лица извършихме реанимиране (resuscitation) на латентни микробни клетки. Добивът на секвенции от секвенирането беше 11,22 Gb, като преминалите през филтъра кълстери от 18,7 МВ (82,7%) и $Q30 > 70\%$. Общият брой на идентифицираните бактериални OTU беше 3890 с общо 829 984 секвенции. За гъбичките получихме 34 980 OTU, състоящи се от 3 046 992 секвенци. Броят на неидентифицираните бактериални таксони беше 13%, а при гъбичките - 17,2%.

Доминиращите бактериални типове сред некултивираните проби бяха Proteobacteria 92,97%, Firmicutes 2,18%, Actinobacteria 1,74% и Planctomycetes 1,55%, докато сред култивираните проби Proteobacteria бяха 47,83%, Firmicutes 25,85%, Actinobacteria 16,42%, Bacteroidetes 3,48%, Cyanobacteria 2,74% и Fusobacteria 1,53%. Видове от разреди Bacillales, Lactobacillales и Corynebacteriales показаха най-добра култивируемост. За бактериалните разреди Sphingomonadales, Rhizobiales и Rhodospirillales доказахме, че остават некултивируеми. Броят секвенции беше равен или дори намален в сравнение с получените брой секвенции от некултивирана кръв. Гъбичките от типове Basidiomycota, Ascomycota и неидентифицираните гъби бяха съответно 65,08%, 17,72% и 17,2% сред некултивираните проби, докато сред култивираните проби те бяха съответно 58,08%, 21,72% и 20,2%. При гъбичките, разреди Tremellales, Filobasidiales и Russulales бяха идентифицирани като некултивируеми. Видовете гъби от разред Pleosporales, Saccharomycetales и Helotiales бяха сред добре култивиращите се. Фигури 6 и 7 показват, че комбинацията от стресови фактори, индуцират механизми за оцеляване на микробната флора, чрез експлозивен растеж. Повечето микробни видове пролиферират добре в тестваната течна ВНИ хранителна среда и условия на култивиране (Tsafarova et al., 2023). Това се вижда от увеличения брой секвенции на OTU след култивиране. Култивираните проби показват значителен брой таксони, които не се откриват в некултивираните. Това доказва необходимостта при

описване на пълното биоразнообразие в кръвта да се тестват двойки култивирани и некултивирани проби.

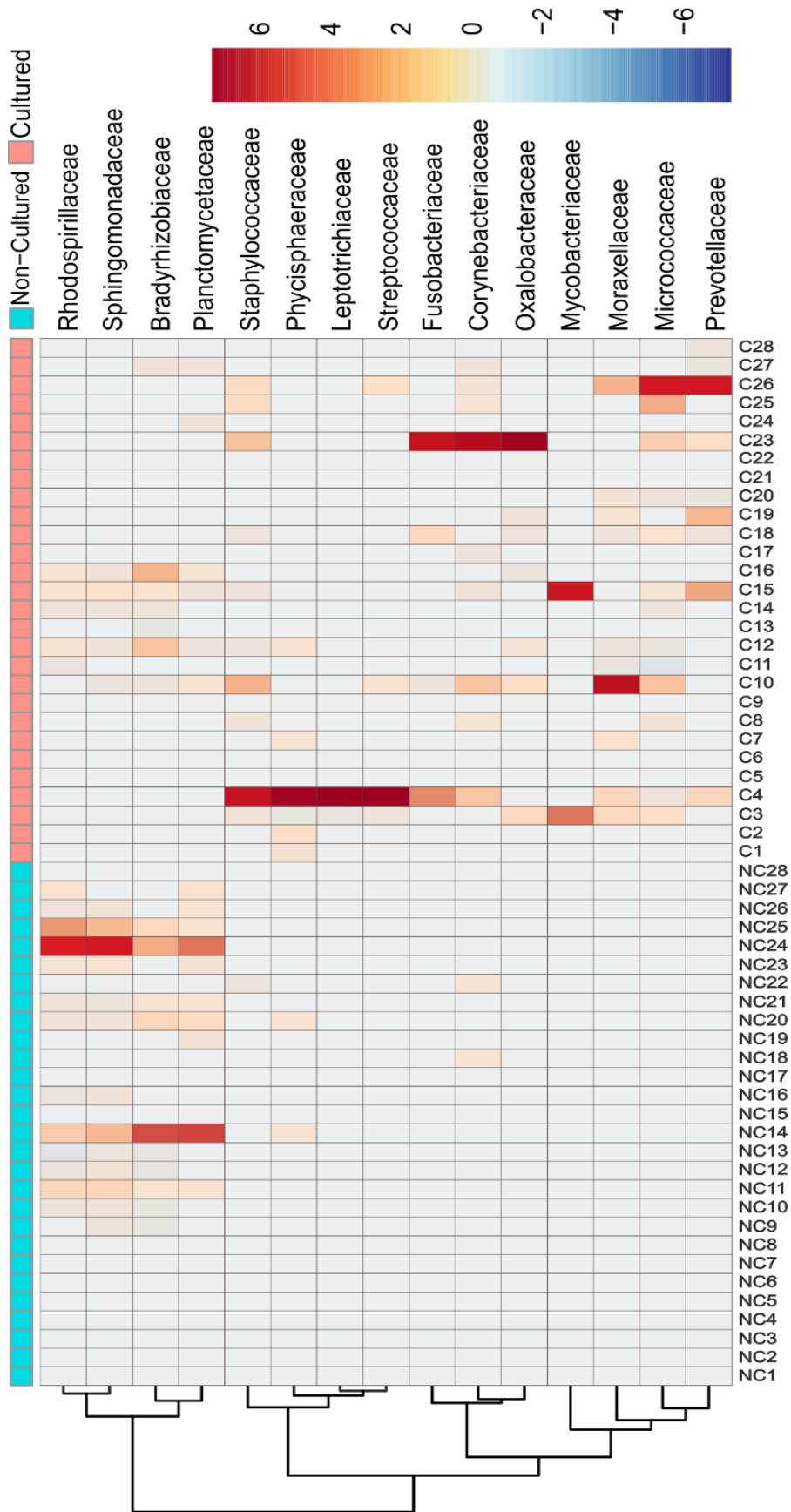
Order	Difference	Family	Difference	Genus	Difference
Bacillales	7040	Staphylococcaceae	6430	Staphylococcus	6430
Lactobacillales	5014	Corynebacteriaceae	4848	Micrococcus	3521
Corynebacteriales	4960	Micrococcaceae	4474	Corynebacterium	2813
Micrococcales	4474	Moraxellaceae	2272	Acinetobacter	2272
Burkholderiales	2975	Streptococcaceae	2105	Streptococcus	2102
Clostridiales	2772	Fusobacteriaceae	1806	Fusobacterium	1806
Pseudomonadales	2627	Oxalobacteraceae	1483	Pelomonas	748
Fusobacteriales	2257	Prevotellaceae	1187	Rothia	597
Bacteroidales	1635	Planctomycetaceae	-434	Brevundimonas	409
Rhodospirillales	-1815	Rhodospirillaceae	-1193	Novosphingobium	-1219
Rhizobiales	-8680	Bradyrhizobiaceae	-8580	Bradyrhizobium	-8487
Sphingomonadales	-14719	Sphingomonadaceae	-14719	Sphingomonas	-13594

Фиг. 6. Разлики в многообразието на идентифицираните бактериални таксони сред некултивирани и култивирани кръвни проби. Броят на OTU секвенциите е показател за ефекта на стресовите условия върху микробната пролиферация.

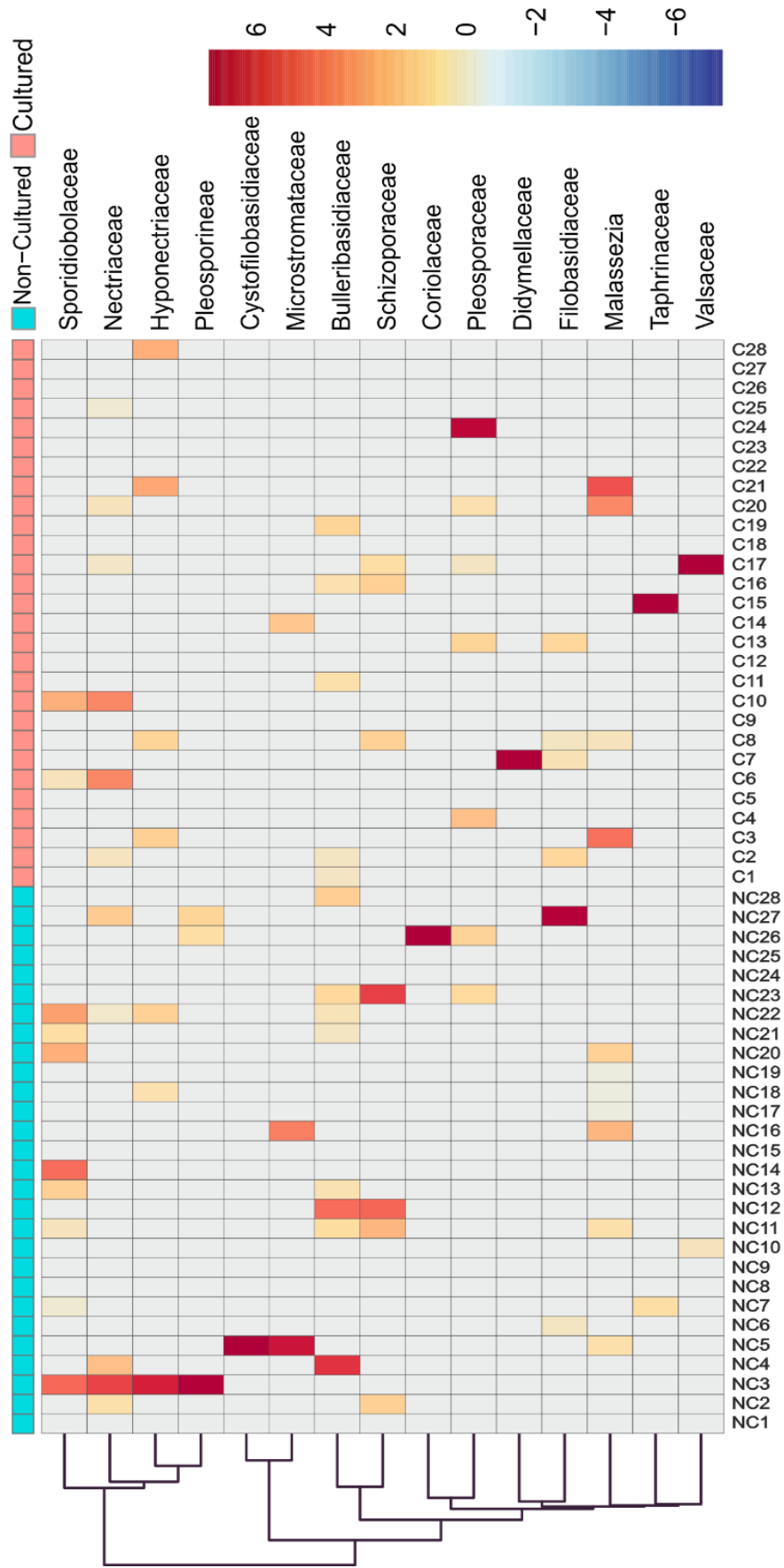
Order	Difference	Family	Difference	Genus	Difference
Pleosporales	81824	Valsaceae	9218	Candida	20443
Saccharomycetales	22346	Pleosporaceae	7447	Ciboria	15530
Helotiales	17259	Taphrinaceae	7277	Taphrina	7277
Diaporthales	7981	Malassezia	7262	Leptosphaeriaceae	7197
Taphrinales	7277	Hyponectriaceae	5731	Alternaria	5990
Malasseziaceae	7262	Pleosporineae	5156	Cytospora	5864
Hypocreales	6705	Nectriaceae	4951	Monographella	5731
Corticiales	-2481	Microstromataceae	-2836	Sporobolomyces	-4384
Microstromatales	-2836	Schizoporaceae	-3405	Perusta	-6916
Sporidiobolales	-4384	Corioliaceae	-3975	Corioliopsis	-9948
Russulales	-4678	Sporidiobolaceae	-4384	Vishniacozyma	-12041
Filobasidiales	-13154	Filobasidiaceae	11026	Dioszegia	-12761
Tremellales	-23731	Bulleribasidiaceae	27167	Filobasidium	-13908

Фиг. 7. Разлики в многообразието на идентифицираните гъбични таксони сред некултивирани и култивирани кръвни проби. Ефектът на стресовите условия върху микробната пролиферация в кръвта е оценен според броя на секвенциите от дадена OTU.

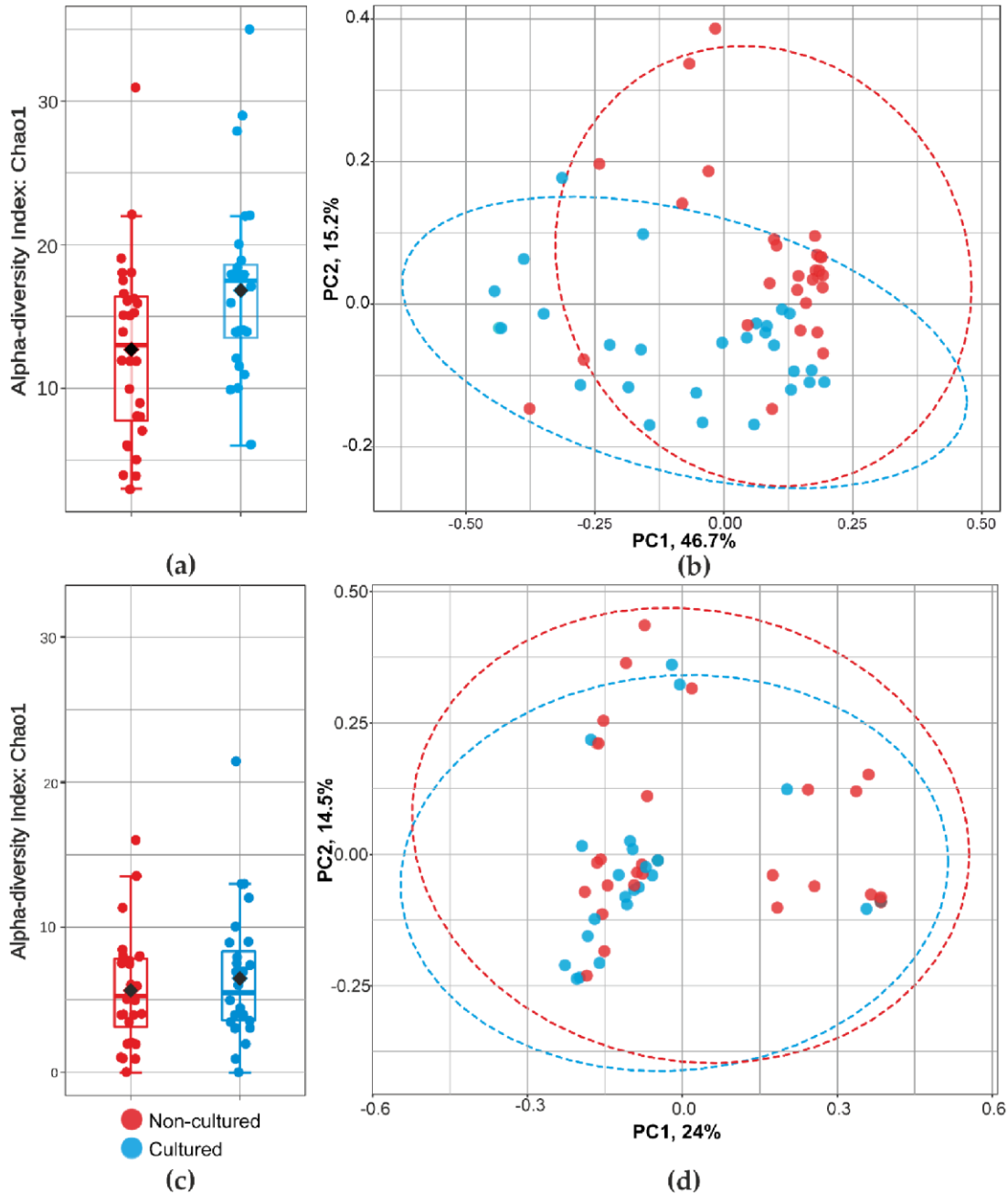
Комбинирайки OTU от всички култивирани и некултивирани проби, бяха идентифицирани общо 105 бактериални рода. Подобни резултати са докладвани за изследвани проби от горния женски репродуктивен тракт, считани за стерилни. Както беше споменато по-горе, на ниво тип установихме, че в микробиома на кръвта преобладава тип Proteobacteria 93%, следван от Firmicutes 2,18%, Actinobacteria 1,74% и Planctomycetes 1,55%. Значително скрито биоразнообразие може да бъде показано в култивирани проби. В изследваните от нас проби установихме, че уникални бяха общо 55 бактериални и 39 гъбични рода, т.е. идентифицират се само в култивираните или само в некултивираните проби.



Фиг. 8. Групиране на бактериалното разнообразие на ниво семейство за некултивирани и култивирани кръвни проби от лица от 1 до 28.



Фиг. 9. Групиране на гъбичното разнообразие на таксономично ниво семейство за некултивирани и култивирани кръвни проби от лица от 1 до 28



Фиг. 10. Анализ на бактериалното и гъбично разнообразие сред култивирана и некултивирана кръвна микрофлора на ниво семейство. Алфа-разнообразие (индекс Chao 1) за бактерии (A) и гъбички (B). Графиките (C) и (D) показват дистанцията между бактериалното и гъбичното бета-разнообразие на култивирани и некултивирани кръвни проби.

4.2. Определяне на микробното разнообразие при саркоидоза

4.2.1 Визуализация на таксономичния състав на единични и групови кръвни микробиоми. Причинно-следствени връзки между дисбиоза на микробиома на кръвта и патогенезата на белодробната саркоидозата.

За двугодишен период от ноември 2020 г. до ноември 2022 г. са взети кръвни проби от 7 пациенти със саркоидоза и 22 здрави индивида. Пробите бяха секвенирани чрез 16S метагеномен подход. След качествен контрол и обработка са получени общо 294 736 сурови секвенции. Броят на бактериалните секвенции в пробите на саркоидозно болните беше значително по-малък и беше нормализиран до 900 секвенции на проба. Таксономичният анализ идентифицира общо 1541 OTU, като 118 OTU имаха праг >10 секвенции. Таблица 2 показва информация за разпределението на OTU секвенциите между ключовите таксони.

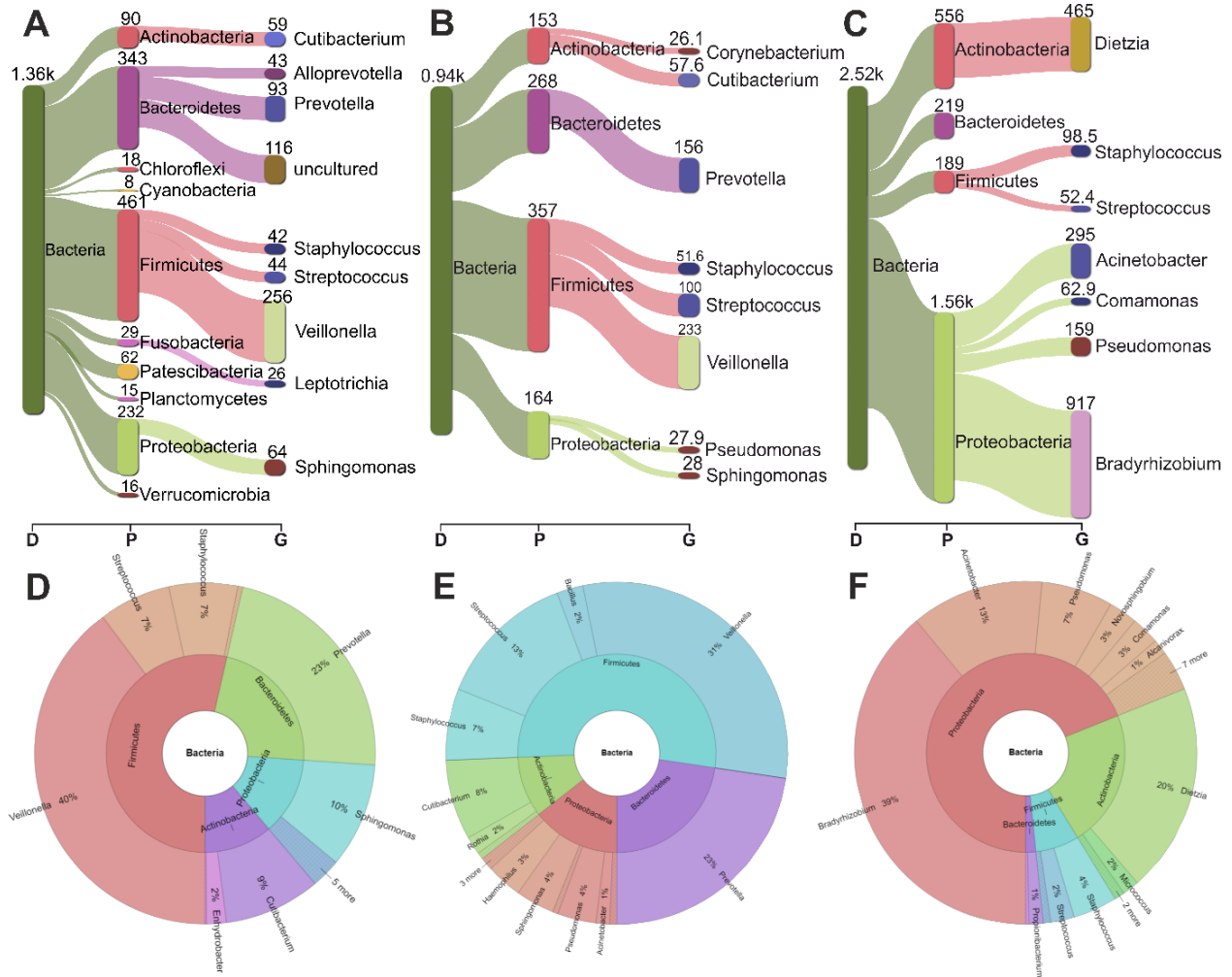
Таблица 2. Общ брой идентифицирани бактериални таксони и тяхното разпределение при пациенти със саркоидоза и здрави контроли.

Таксон	Контролна група	Саркоидоза
Тип	13	13
Семейството	70	70
Род	118	102
Общ брой секвенции	55451	6558
Среден брой секвенции на проба (\pm SD)	2520 (\pm 293)	937 (\pm 58)

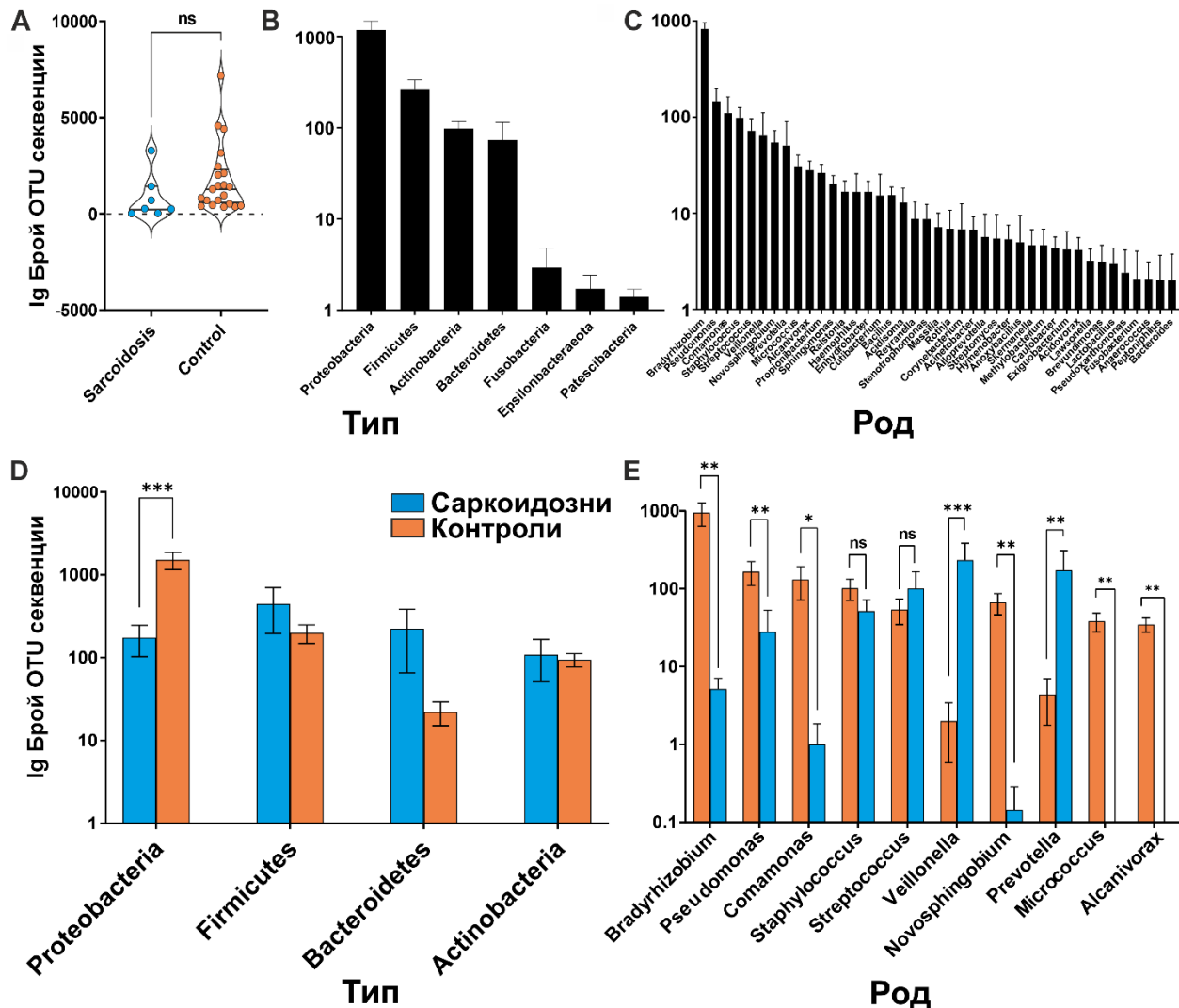
SD – стандартно отклонение

Данните за микробиома на всеки пациент със саркоидоза бяха визуално представени на ниво тип и род чрез два алтернативни графични подхода – диаграма на Sankey (фиг. 11 A,B,C) и концентрична кръгова диаграма тип Krona (фиг. 11 D,E,F). На фигури 11A и 11D е анализът на единичен пациент със саркоидоза. Графиката описва брой секвенции за таксон. На фигури 11B и 11E е показано разнообразието на кръвния микробиом, представен от средния брой бактериални секвенции за таксон в саркоидозната група, докато фигури 11C и 11F изобразяват същия анализ за групата здрави индивиди. В диаграмите тип Krona,

относителният дял на бактериалните секвенции за таксон е представен в проценти. Тези методи дават възможност за лесно интерпретиране на резултатите, като се използва цветно визуално описание на таксономичното съдържание на ниво тип и род. Таксономичното разнообразие е представено от дебелината на клоните в диаграмата на Sankey и ъгловия сектор в кръговата диаграма тип Krona.



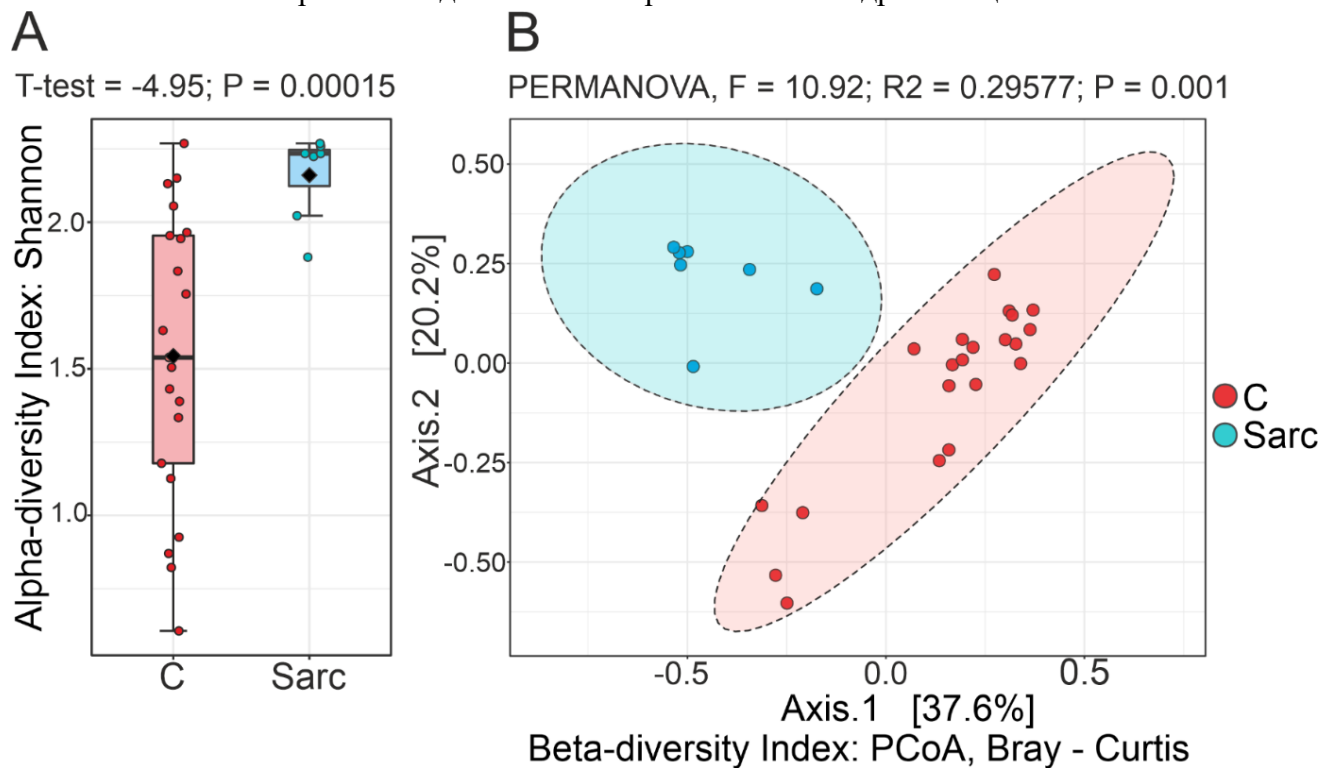
Фиг. 11. Бактериален състав. Визуализация на индивидуалния и групов кръвен микробиом на ниво тип и род представени чрез диаграми на Sankey и Krona. (A и D) единичен пациент с белодробна саркоидоза (Sarcl); (B и E) средногрупова стойност на броя секвенции за таксон при групата от саркоидозни пациенти; (C и F) средногрупова стойност на броя секвенции за таксон при групата от здрави лица. (таксономична скала: D – домейн, P – тип, G – род).



Фиг. 12. Количествено описание на микробиомното разнообразие на родово и типово ниво. А. Общ брой идентифицирани секвенции за всяка проба и група, представени в десетична логаритмична скала. В. Седем идентифицирани типа в кръвта на изследваните лица и С. 41 рода. D. Средна групова разлика между брой идентифицирани секвенции представени в десетична логаритмична скала при контролната и саркоидозна групи според количественото представяне на индивидуалните таксони на ниво тип, D и род, E. Значимостта на груповата статистика изчислена чрез теста на *Kruskal-Wallis*: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; нз - не е значимо; SEM - Standard Error of the Mean, стандартна грешка.

Диаграмата на Sankey показва действителното разнообразие от секвенции, а Krona илюстрира относителното богатство. И двете диаграми показват четири доминантни типа: Actinobacteria, Proteobacteria, Firmicutes и Bacteroidetes. Микробиомният анализ на саркоидозно болните идентифицира няколко специфични родове - *Cutibacterium*, *Corynebacterium* и *Veillonella*. Род *Dietzia* и *Acinetobacter* преобладават сред здравите лица, докато род *Streptococcus* присъства и в двете групи с леко увеличение, наблюдавано в групата на саркоидозно болните.

Данните за микробиомния състав в кръвта на 7 пациента с белодробна саркоидоза бяха статистически сравнени с данните за микробиома на 22 здрави лица.

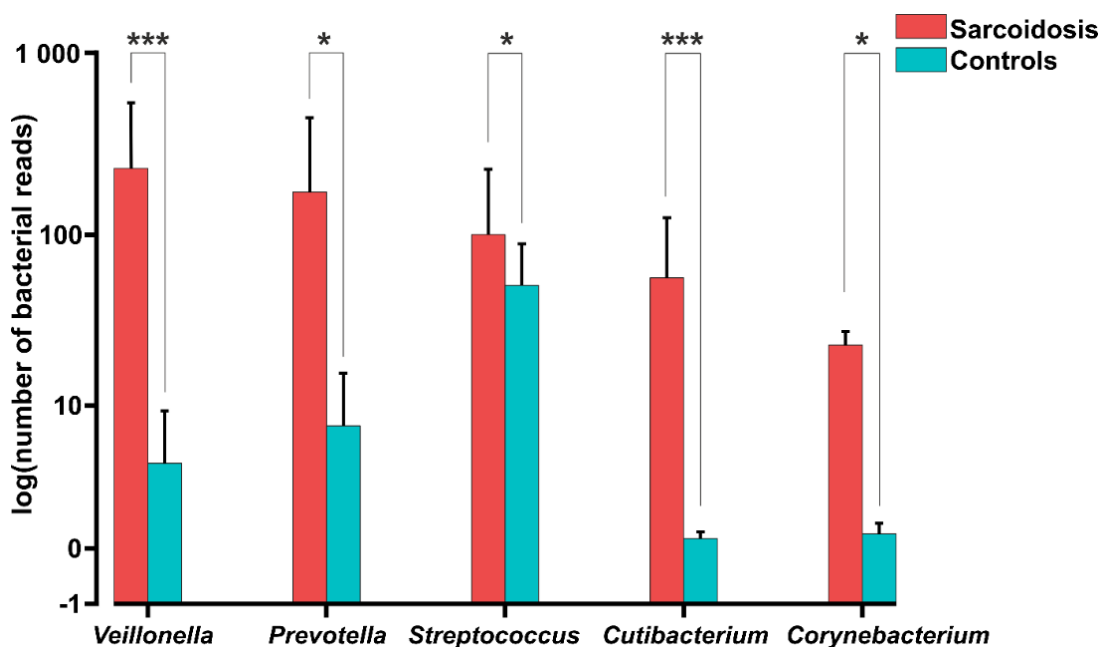


Фиг. 13. Статистическо сравнение на състава на кръвните микробиоми между здрави лица и саркоидозно болни. (А) Алфа разнообразие – медиана на индекса на Shannon в контролната група и групата на саркоидозно болни. (Б) Сравнение на таксономичния състав на микробиома на двете групи изчислен чрез анализ на главните координати (PCoA), използвайки Bray - Curtis бета-индекс, (съкращения: C – контроли, Sarc – саркоидоза).

Статистическият анализ показва значително увеличение на богатството на бактериалното алфа-разнообразие при пациенти със саркоидоза, използвайки индекса на

Shannon в сравнение със здрави контроли ($2,39 \pm 0,08$ спрямо $1,57 \pm 0,63$; $F(1,28) = 4,95$; $P=0,00015$) (фиг. 13А). Индексът на Bray-Curtis за различие в таксономичния състав беше използван, като показател за бета-разнообразие между групите на пациенти и здрави лица. РСоА анализът ясно илюстрира разграничаването между двете групи (Фиг. 13В).

Прилагайки теста на Mann-Whitney ние установихме значително увеличение на пет бактериални рода при пациенти със саркоидоза: *Veillonella*, *Prevotella*, *Cutibacterium*, *Corynebacterium* и *Streptococcus* (Фиг. 14).



Фиг. 14. Логаритмично преобразувано средно микробно разнообразие по Shannon на пет рода *Veillonella*, *Prevotella*, *Cutibacterium*, *Corynebacterium* и *Streptococcus*. (\pm стандартна грешка от средната стойност)

4.3. Машинно обучение. Дърво на решенията и патогенеза на саркоидозата

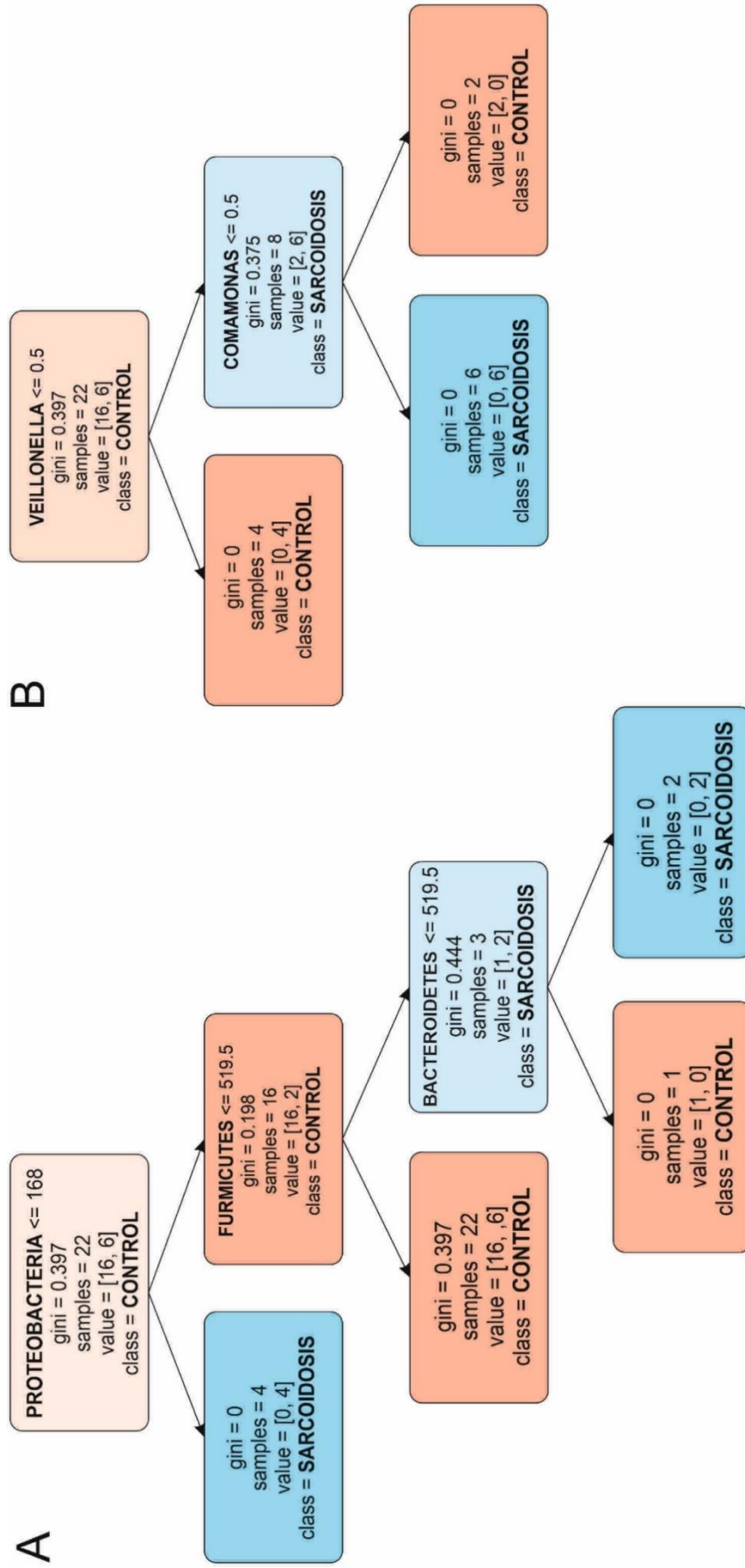
4.3.1. Начален биостатистически анализ

4.3.2. Машинно обучение. Класификационно дърво на решенията

Беше създаден класификационен модел на йерархично бинарно дърво на решенията с алтернативни класове „контроли“ и „саркоидозно болни“ (Фиг. 15). Класификацията беше

направена на базата на количеството OTU секвенции за таксон. Фигура 24 А показва структурата на дървото, изчислена от данните за многообразието на таксономичното ниво „тип“. Правоъгълниците представляват класификационни възли. Всички възли съдържат информация за процеса на класификация: праг, брой на представителите, коефициент на грешките (Gini) и дял на вярно спрямо грешно класифицирани проби. От всеки възел, съдържащ представители и на двете групи за класифициране (болни и здрави лица) произлиза разклонение, показано със стрелки. Процесът на класификация продължава, докато се достигнат възли, които съдържат представители само от един клас. Те не се разклоняват повече и се означават като терминални. Силата на яркостта на цвета на всеки възел показва точността/прецизността на класификацията.

При нулевия (основен) възел, тип *Proteobacteria* диференцира контролната от саркоидозна групи при праг 168 OTU секвенции. Началният възел съдържа общо 22 проби, включващи 16 контроли и 6 пациента със саркоидоза, като коефициентът за грешка е $Gini = 0.397$. Съответният ляв възел (възел 1), произлизащ от началния, съдържа четири проби, всички правилно класифицирани като саркоидозни ($Gini = 0$). Следователно възел 1 е краен (терминален) възел и не се разклонява повече. Класификацията в десният възел 2 се основава на броя OTU секвенции на тип *Firmicutes* с праг от 519,5. Той съдържа общо 18 проби и е класифициран като принадлежащ към контролната група (16 контроли спрямо 2 саркоидози; $Gini = 0.198$). Съответният десен възел 3 е краен възел с 22 проби и се класифицира като контролен възел (16 контролни проби спрямо 6 проби саркоидоза; $Gini = 0.397$). Възел 4 е разклонен от тип *Bacteroidetes* при праг на многообразие 519,5 ($Gini = 0,444$). Възел 4 е класифициран като саркоидозен възел съдържащ 2 саркоидозни спрямо 1 контролна проба. По-нататък възел 4 се разделя на възли 5 и 6, които са крайни възли със стойност $Gini = 0$ (Фигура 15 А). Машинното обучение проведено върху данните за таксономичния състав на ниво тип показва, че са необходими четири нива на класификация и три таксономични типа за пълното разграничаване между саркоидозните и контролни проби. В случая, водещо значение имаше тип *Proteobacteria*, следван от типове *Firmicutes* и *Bacteroidetes*.



Фиг. 15. Диаграми, представящи класификационни дървета. А. таксономично ниво тип и В. таксономично ниво род. Възлите, класифицирани като саркоидозни са означени в синьо, а контролните в оранжево

Трябва да се подчертае, че прилагайки алгоритъма на класификационното дърво на решенията ние изчислихме приноса на всеки микробен тип за диференциране на саркоидозно болните лица от здрави лица. Proteobacteria се откроява като най-влиятелният тип в класификационния модел (0,593), следван от Firmicutes (0,255) и Bacteroidetes (0,153). Представителите на тип Actinobacteria не показват принос към модела (0).

Фигура 15 В показва структурата на дървото на решенията на ниво род. Класификацията започва с основен възел, който се разделя на базата на род *Veillonella*, при праг от 0.5; Gini = 0.397. Съдържа 22 проби и е класифициран като контролен възел (16 контролни спрямо 6 саркоидоза). След това левият възел (възел 1), произлизащ от основния, съдържа проби, при които OTU секвенциите на *Veillonella* са по-малко или равни на 0,5. Възел 1 е терминален възел, съдържащ 4 проби на здрави лица (Gini = 0). Десният възел (възел 2) съдържа проби, при които *Veillonella* надвишава 0,5 и се разклонява допълнително на базата на OTU секвенции на род *Comamonas*. Разделянето на възел 3 се определя с праг от 20 OTU секвенции и се характеризира с коефициент на Gini = 0.375. Класифицира се като саркоидоза и съдържа 2 контроли спрямо 6 саркоидозни проби. Този възел се разделя на два крайни възела – възел 4 и възел 5, като и двата показват индекс на Gini = 0.3750. Възел 4 е класифициран като саркоидоза с 6 проби от саркоидозно болни, докато възел 5 съдържа 2 контролни проби.

Особеното значение за класификацията на групите пациенти и здрави лица на ниво род има род *Veillonella* с принос 0,656, следван от род *Comamonas* с принос 0,344. За разлика от тях родовете *Bradyrhizobium*, *Pseudomonas*, *Staphylococcus*, *Streptococcus*, *Novosphingobium*, *Prevotella*, *Micrococcus* и *Alcanivorax* показват липса на принос в класификацията. За всеки от тези родове е изчислен принос към класификацията равен на нула.

За трениране и валидиране на модела на дървото на решенията беше взета на случаен принцип подгрупа от общия брой проби, равни на 7 проби (6 контролни и една саркоидозна), и бяха тествани на ниво тип и род. Резултатите от тази оценка бяха идентични с гореописаните. Всички 7 изследвани лица бяха класифицирани правилно със стойности на фалшиво положителни и фалшиво отрицателни класификации равни на 0.

5. Заключение

5.1. Състав на кръвния микробиом в норма

Нашите резултати показват, че кръвната микрофлора е жизнеспособна екологична ниша. Кръвните проби, култивирани при стресови условия, показаха скрито микробно разнообразие, което в некултивирани проби може да остане неоткрито дори при метагеномен анализ. Чрез култивиране на кръвната микробиота, ние идентифицирахме добре култивируеми бактерии и гъбички. Нашите резултати добре корелират с наскоро получени резултати за човешкия кръвен микробим, където чрез култивиране и метагеномен анализ се доказва присъствие на множество микробни видове в кръвта на здрави лица. Всички изследвания цитират сходен микробиомен състав от Proteobacteria, Actinobacteria, Firmicutes и Bacteroidetes. Нашият колектив е все още единственият, който докладва резултати за гъбичния микробиом в кръвта на здрави лица. Изключение правят изследвания на тъканния микробиом при пациенти с различни онкологични заболявания, където се доказват бактерии и гъбички.

5.2. Състав на кръвния микробиом и визуализация на данните при саркоидоза

Кръвният микробиом досега не е изследван при пациенти със саркоидоза. Ние изследвахме и сравнихме кръвния микробиомен състав чрез 16S таргетно секвениране при 7 пациенти със саркоидоза и 22 здрави контроли. Седемте пациенти бяха диагностицирани за първи път и не са били лекувани за саркоидоза с противовъзпалителни средства, които потенциално биха могли да повлияят на състава на микробиома. Целта на изследването беше да идентифицираме специфични микробни таксони, характерни при саркоидозата. Ние доказахме статистически значими увеличения на определени микробни таксони при саркоидозни пациенти сравнено със здрави лица.

Бяха приложени диаграми на Sankey и кръгови диаграми тип Krona за визуализиране на микробиомни данни за единични лица или малки групи. Приложихме двата метода, за да илюстрираме таксономичния състав при единични пациенти и малка група саркоидозно болни, които сравнихме с единични и групови диаграми за здрави контроли (Фиг. 11). Двата метода демонстрираха многообразието на микробиомния състав и относителното разнообразие на основните типове Actinobacteria, Proteobacteria, Firmicutes и Bacteroidetes. Диаграмата на Sankey е по-рядко използвана за визуализация на данни сравнено с кръговите

диаграми Krona. Нашето мнение е, че Sankey диаграмата представя резултатите по-нагледно и разбираемо. Тя по подходящ интуитивен начин дава възможност на неспециалисти да се ориентират в количествения и качествен микробен състав на пробите. Методите за визуализация на микробиома могат да бъдат полезни в клиничната практика, като подпомагат диагностиката, проследяват прогресията на заболяването, дават възможност за персонализирана медицина, идентифицират потенциални патогени и улесняват научните изследвания. По този начин могат да се извлекат максимално ценни клинични изводи от данните за микробиома на единична или малка група пациенти.

Използвахме алфа и бета показатели за разнообразие и параметрични и непараметрични методи за статистически извод, за да определим микробния състав в кръвта на пациенти с белодробна саркоидоза и здрави лица. Пет рода, *Veillonella*, *Prevotella*, *Cutibacterium*, *Corynebacterium* и *Streptococcus*, показаха статистически значимо количествено увеличение в разнообразието при пациентите със саркоидоза в сравнение с контролната група. Петте рода се състоят от човешки коменсални или симбиотични видове, както и патогени или видове с патогенен потенциал. Анализът на данните установи, че доказаните родове могат да се ползват като потенциални микробни маркери за саркоидоза.

Мониторингът на таксономичния състав на микробиома на кръвта дава възможност за оценка на дисбиозата, свързана със саркоидоза. Микробната дисбиоза може да бъде описана от микробните типове или родове, които са обогатени в резултат на патогенезата на заболяването. Нашите изследвания косвено се потвърдиха от публикувани метагеномни анализи на БАЛ и бронхиална тъканна биопсия, където е идентифициран сходен микробиомен състав при пациенти със саркоидоза. Zimmermann et al. идентифицират в БАЛ родовете *Atopobium* и *Fusobacterium* като потенциални етиологични причинители на саркоидоза. Gupta et al. провежда сравнителен анализ на алвеоларния микробиом при различни респираторни заболявания ХОББ, ИББ и саркоидоза като доказва, че Actinobacteria и Proteobacteria са значително увеличени при пациенти със саркоидоза. Ние наблюдавахме по-слабо присъствие на Actinobacteria и Proteobacteria, докато Firmicutes бяха количествено увеличени на индивидуално и групово ниво (Фиг. 11 и 14). Микробното алфа разнообразие е значително по-високо в кръвните проби от саркоидозно болни в

сравнение със здрави лица (Фиг. 13 А). Разликите в микробния състав бяха потвърдени от увеличения индекс за бета разнообразие, показано на фигура Фиг. 13 В.

5.3. Патогенен потенциал на доказаните родове

Cutibacterium acnes е коменсал на кожата. Разпространен е повсеместно сред здрави индивиди. *Cutibacterium acnes* произвежда липази, които причиняват възпаление на кожата и други части на тялото. *Cutibacterium acnes* е открит в грануломна тъкан чрез имунохистохимия и се смята, че участва в патогенезата на саркоидозата. Чрез биоинформационен сравнителен анализ са идентифицирани генни хомологии между *C. acnes* и други микробни видове, които имат потенциал за образуване на гранулом, като *Yersinia pseudotuberculosis* и *Mycobacterium spp.*.

Streptococcus spp. са идентифицирани при пациенти със саркоидоза. Сред стрептококите най-патогенни са *S. pneumoniae* и *S. pyogenes*, но са докладвани и коменсали с патогенен потенциал. Проучванията показват, че някои щамове на *Streptococcus* могат да допринесат за образуването на гранулом чрез активиране на имунната система. Установено е, че *Streptococcus spp.* може да активира Toll-like рецепторите (TLR), което води до производство на провъзпалителни цитокини.

Видове от род *Corynebacterium*, особено *C. diphtheria*, са белодробни патогени, които причиняват възпаление и предизвикват имуен отговор. Проучванията показват, че *Corynebacterium spp.* може да индуцира образуването на гранулом чрез активиране на NLRP3 инфламазома, което води до синтез на провъзпалителни цитокини. *Streptococcus spp.* и *Corynebacterium spp.* не са етиологични причинители на саркоидоза, а по-скоро са придружаващи белодробни инфекции. Това схващане се подкрепя от наблюдаваното им увеличение при белодробни заболявания, различни от саркоидоза.

Prevotella spp. и *Veillonella spp.* са предимно коменсални видове, които са признати за опортюнистични патогени, свързани с чревни и орални инфекции. Освен това тези родове са открити в нормалната белодробна микрофлора. Двата рода имат обогатено присъствие в БАЛ на пациенти със саркоидоза. Ролята на *Veillonella* при образуването на грануломи все още не е напълно изяснена. Видовете *Veillonella*, обаче синтезират млечни и други мастни киселини, променяйки рН на локалната среда причинявайки възпалителен отговор, водещ

до образуване на гранулом. Някои видове *Prevotella* могат да участват в етиологията на саркоидозата чрез активиране на имунната система. Проучванията показват, че видовете *Prevotella* в чревния микробиом корелират с повишени нива на провъзпалителни цитокини.

5.3. Машинно обучение. Класификационно дърво на решенията

Описаният и приложен анализ на класификационното дърво на решенията успешно разграничи саркоидозно болните от контролната група здрави лица. Използвахме данни от предишните експерименти с цел да допълним и разширим информацията за потенциалния принос на описаните микробни типове и родове в патогенезата на саркоидозата. Визуализацията на дървото на решенията даде представа за процеса на вземане на решения. Представителите на тип Proteobacteria имаха най-голямо значение за класификацията (с коефициент на принос = 0.593). Интересното е, че тип Firmicutes (0.255) имаше почти удвоен коефициент в сравнение с тип Bacteroidetes (0.153), което предполага, че промените в неговото разнообразие могат да бъдат решаващ показател за дисбиозата на микробиома в кръвта, свързана със саркоидоза. Приносът на актинобактериите в класификационния процес беше равен на нула, което предполага, понижена вероятност за участие в етиологията на саркоидозата.

На ниво род, родът *Veillonella* имаше най-голямо значение за класификацията (0.656), което показва, че той играе важна роля. Родът *Comamonas* е втори по важност в класификацията (0.344). Според литературата са открити щамове *Comamonas*, идентифицирани при пациенти с кистозна фиброза (муковисцидоза) на дихателната система. Останалите родове не допринасят към модела за класификация и нямат принос към патогенезата на саркоидозата.

Добре изразените бифуркации в дървото на решенията, основани на тези родове, допълнително подчертават тяхното потенциално синергично значение в патогенезата на белодробната саркоидоза. Въпреки това, прякото участие на *Comamonas* и *Veillonella* в образуването на грануломи при грануломатозните белодробни болести, се нуждае от допълнително експериментално потвърждение.

6. Изводи

Многобройни проучвания доказват наличието на микробни видове в тъканите, кръвта и телесните течности при болни и здрави хора. Ролята на кръвния микробиом в човешкото здраве е отворено поле за иновативни изследвания с потенциал да се превърне в нова дисциплина на медицинската микробиология. Може да се предположи, че през следващите години нови изследователски екипи ще се съсредоточат върху изучаването на микробиома на кръвта и връзката му със заболяванията с предполагаема инфекциозна етиология, като артрит, саркоидоза, кръвна анемия, латентна туберкулоза и много други (Tsafarova, 2021).

Изводите от проведените изследвания и получените резултати може да се обобщят в следното:

1. Приложените стресови условия и хранителна среда са подходящи за реанимиране на кръвните микробиоти. Метагеномният анализ дава възможност, според броя на секвенциите за таксон, количествено да се определи култивируемостта на кръвените микробиоти.

2. В култивирани и некултивирани кръвни проби доказахме 241 таксономични единици (OUT) разпределени в 40 бактериални разреда, 66 семейства и 105 рода. Некултивируемите бактериални разреда бяха Sphingomonadales, Rhizobiales и Rhodospirillales. Най-добре култивируеми в използваната от нас хранителна среда и стресови условия бяха представителите на разреда Bacillales, Lactobacillales и Corynebacteriales.

3. Доказахме, че кръвния микробиом съдържа 272 гъбични таксономични единици, разпределени в 61 разреда, 105 семейства и 133 рода. Некултивируемите гъбични разреда бяха Tremellales, Polyporales и Filobasidiales. Най-добре култивируеми в използваната от нас хранителна среда и стресови условия бяха гъбички от разреда Pleosporales, Saccharomycetales и Helotiales.

4. Доминиращите бактериални таксономични типове в некултивирана цяла кръв бяха Proteobacteria - 92.97%, Firmicutes - 2.18%, Actinobacteria - 1.74% и Planctomycetes - 1.55%, докато в култивираните кръвни проби доминираха Proteobacteria - 47.83%, Firmicutes

- 25.85%, Actinobacteria - 16.42%, Bacteroidetes - 3.48%, Cyanobacteria - 2.74% и Fusobacteria - 1.53%.

5. Доминиращите гъбични таксономични типове в некултивирана цяла кръв бяха Basidiomycota 65.08%, Ascomycota 17.72%, и неидентифицирани гъбички 17.2%. В култивираните кръвни проби преобладаваха същите гъбични типове, но с поменено процентно съотношение, съответно 58.08%, 21.72% и 20.2%.

6. Диаграмите на Sankey и Krona ефективно илюстрират таксономичното разнообразие на кръвните микробиоми на индивидуално и групово ниво. Според нашия опит диаграмата на Sankey е визуално по-интуитивна и по-удобна за сравнения на няколко таксономични нива едновременно. Приложенията от нас подходи за визуализиране на метагеномни данни имат практическа стойност за описание на микробиома при редки болести, като саркоидоза.

7. Параметричните статистически методи доказаха повишение на средния брой на секвенциите на родове *Veillonella*, *Prevotella*, *Cutibacterium*, *Corynebacterium* и *Streptococcus* при саркоидозно болни в сравнение със здрави лица.

8. Приложеният от нас класификационен модел за машинно обучение идентифицира таксоните с най-голям възможен принос за патогенезата на саркоидозата. На базата на ампликонен метагеномен анализ, ние идентифицирахме, че типове Proteobacteria, Firmicutes и Bacteroidetes и родове *Comamonas* и *Veillonella* при саркоидозно болни имат най-висок изчислен коефициент на принос за диференциране на саркоидозни от клинично здрави лица.

7. Приноси

7.1. Фундаментални

7.1.1. За пръв път са демонстрирани експериментални доказателства за връзка между таксономичния състав на кръвния микробиом и развитието на белодробна саркоидоза.

7.1.2. Предложена е хипотеза за взаимодействие между кръвния и белодробния микробиом в патогенезата на саркоидозата.

7.1.3. За първи път, културелно и метагеномно се доказва присъствието на гъбичен микробиом в кръвта при здрави лица.

7.1.4. Количествената оценка на дисбиозата на кръвния микробиом е показател за развитие на патологични процеси.

7.2. Приложни

7.2.1. Разработен е биоинформатичен подход за установяване на състава на кръвния микробиом при здрави възрастни лица.

7.2.2. Предложени са методи за визуализация и оценка на количествения и качествен състав на кръвния микробиом на единични пациенти.

7.2.3. Разработен е ефективен биоинформатичен метод за сравнителен анализ на таксономичния състав на култивирани кръвни проби за изпитване на различни хранителни среди и стресови условия за реанимиране на латентните кръвни микробиоти.

7.2.4. Приложеният класификационен модел за машинно обучение дава възможност за определяне на ключовите микробни таксони с най-голям вероятен принос в патогенезата на различни заболявания свързани с микробна дисбиоза.

8. Публикации

8.1. Публикации в научни журнали по темата на дисертационният труд

1. Panaiotov, S, **Hodzhev, Y**, Tsafarova, B, Tolchkov, V, Kalfin, R. (2021). Culturable and non-culturable blood microbiota of healthy individuals. *Microorganisms*, 9(7), 1464.
ИФ 4.128, H-17, SJR 0.91, Q2, (11 цит.)
2. **Hodzhev, Y**, Tsafarova B, Tolchkov V, Youroukova V, Ivanova S, Kostadinov D, Yanev N, S. Panaiotov. (2023). Visualization of the individual blood microbiome to study the etiology of sarcoidosis. *Computational and Structural Biotechnology Journal* 22: 50-57.
ИФ 6.155, H-index 65, SJR 1.58, Q1
3. **Hodzhev, Y**. (2023). Analysis of blood microbiome dysbiosis in pulmonary sarcoidosis by decision tree model. *Biotechnology & Biotechnological Equipment*, 37(1), 2283133.
IF 1.6 , SJR 0.317, H-index 48, Q3
4. **Hodzhev, Y**, Tsafarova B, Tolchkov V, Youroukova V, Ivanova S, Kostadinov D, Yanev N, S. Panaiotov. (2023). Decoding microbiome dysbiosis through metagenomic alpha diversity. Implications for sarcoidosis etiology. *Problems of Infectious and Parasitic Diseases*, Accepted, in press
H-index 6, Q4
5. Б. Цафарова, С. Иванова, **Й. Ходжев**, В. Толчков, Н. Янев, Д. Костадинов, В. Юрукова, В. Миланов, С. Панайотов. Отрицателни резултати за *Mycobacterium tuberculosis* и *Cutibacterium acnes* при анализ на включени в парафин тъканни проби от пациенти с туберкулоза и саркоидоза – *Обща медицина (General Medicine)*, 24, 2022, № 3, 11-18
SJR 0,1, H-index 3, Q4
6. Tsafarova B, **Hodzhev Y**, Yordanov G, Tolchkov V, Kalfin R and Panaiotov S (2023) Morphology of blood microbiota in healthy individuals assessed by light and electron microscopy. *Front. Cell. Infect. Microbiol.* 12:1091341. doi:10.3389/fcimb.2022.1091341
IF 6.037, H-index 87, SJR 1.39, Q1 (2 цит.)
7. **Hodzhev Y**, Tsafarova B, Tolchkov V, Panaiotov S. Relations between blood and tissue microbiome in sarcoidosis patients. *ERJ Open Res* 8: 215, 2022. doi:10.1183/23120541.LSC-2022.215. постер.
IF 4.6, SJR 1.3, H-index 35, Q1

Цитиран от:

1. Kim J, Lee S, Moodley Y, Yagnik L, Birnie D, Dwivedi G. The role of the host-microbiome and metabolomics in sarcoidosis. *Am J Physiol Cell Physiol.* 2023 Nov 1;325(5):C1336-C1353. doi: 10.1152/ajpcell.00316.2023.
Общо инпакт фактор: 17.959 (22.559)
Брой цитирания: 14

8.2. Участия в конференции и доклади свързани с дисертационния труд

1. B. Tsafarova, **Y. Hodzhev**, V. Tolchkov, N. Yanev, V. Youroukova, S. Ivanova, V. Milanov, D. Kostadinov, S. Panaiotov. Comparison of Blood and Lung Biopsy Core Microbiome Profiles of Patients with Lung Carcinoma. *Eur Respir J* 2022; 60: Suppl. 66, 3438. Sept. 2022, Barcelona, Spain <https://doi.org/10.1183/13993003.congress-2022.3438>
2. **Yordan Hodzhev**, Borislava Tsafarova, Vladimir Tolchkov, Stefan Panaiotov, Relations between blood and tissue microbiome in sarcoidosis patients. ERS Lung Conference, March 2022. Eshtoril, Spain
3. **Yordan Hodzhev**, Borislava Tsafarova, Vladimir Tolchkov, Nikolay Yanev, Natalia Gabrovska, Silvia Ivanova, Dimitar Kostadinov, Stefan Panaiotov. Lung carcinoma microbiome evaluation based on blood and tissue metagenome sequencing. 32nd European Congress of Clinical Microbiology & Infectious Diseases, Lisbon, Portugal from April 23-26 -ECMID International Congress 2022.
4. **Yordan Hodzhev**, Borislava Tsafarova, Vladimir Tolchkov, Nikolay Yanev, Vania Youroukova Silvia Ivanova, Dimitar Kostadinov, Stefan Panaiotov. The Microbiome and the Lung Tumor Growth. Visualizing Biological Data - VIZBI 2022 <https://vizbi.org/Posters/2022/vA05>
5. **Y. Hodzhev**, B. Tsafarova, V. Tolchkov, N. Yanev, V. Youroukova, S. Ivanova, V. Milanov, D. Kostadinov, S. Panaiotov. Metagenomic profile of blood microbiome associated with pulmonary sarcoidosis. ECCMID 2023, Copenhagen, Denmark 15-18.04.23.
6. **Y. Hodzhev**, B. Tsafarova, V. Tolchkov, R. Kalfin, S. Panaiotov. What does the ratio between blood bacterial and fungal microbiomes abundance tell us about fungal-bacterial interactions? Implications for microbiome modulation by gender and blood group biological factors. HCMPH 2023. 5th Annual Harvard Chan Microbiome in Public Health Center symposium. 15.05.23. Boston, USA.
7. **Y. Hodzhev**, B. Tsafarova, V. Tolchkov, N. Yanev, V. Youroukova, S. Ivanova, D. Kostadinov, S Panaiotov. Bronchoalveolar lavage (BAL) and lung tissue biopsy (LTB) microbiomes in sarcoidosis patients. 7th European Congress for Bronchology and Interventional Pulmonology (ECBIP2023) and XXVII National Congress of the Spanish Association of Respiratory Endoscopy and Interventional Pulmonology (AEER2023) May 11 - 13, 2023, Madrid, Spain
8. **Y. Hodzhev**, B. Tsafarova, V. Tolchkov, N. Yanev, V. Youroukova, S. Ivanova, D. Kostadinov, S. Panaiotov. Comparison of bronchoalveolar lavage (BAL) and lung tissue biopsy (LTB) procedures in the evaluation of respiratory tract microbiota in sarcoidosis (ID 101). 4th International World of Microbiome Conference, 26-28 October 2023, Sofia, Bulgaria.
9. **Y. Hodzhev**, B. Tsafarova, V. Tolchkov, S. Panaiotov. Individual and geographic specificities of the human blood microbiome (ID 218). 4th International World of Microbiome Conference, 26-28 October 2023, Sofia, Bulgaria
10. B. Tsafarova, **Y. Hodzhev**, V. Tolchkov, N. Yanev, V. Youroukova, S. Ivanova, D. Kostadinov, S. Panaiotov. Resuscitation of blood microbiota in a sarcoidosis patient confirmed by transmission

electron microscopy and 16S rDNA sequencing (ID 130). 4th International World of Microbiome Conference, 26-28 October 2023, Sofia, Bulgaria

11. **Y. Hodzhev**, B. Tsafarova, V. Tolchkov, R. Kalfin, S. Panaiotov. Bacterial-fungal microbial interactions in the blood: effects of gender and blood type. (oral presentation). 4th International World Of Microbiome Conference. 26-28 October 2023, Sofia, Bulgaria

12. S. Panaiotov, **Y. Hodzhev**. The human blood microbiome: microbial morphology, life cycle and bacterial-fungal interactions. 4th International World of Microbiome Conference, 26-28 October 2023, Sofia, Bulgaria

13. S. Panaiotov, B. Tsafarova, **Y. Hodzhev**, G. Yordanov, V. Tolchkov, R. Kalfin. The human blood microbiome: microbial metagenomics, morphology and life cycle. 4th International World of Microbiome Conference, 26-28 October 2023, Sofia, Bulgaria

14. S. Panaiotov, B. Tsafarova, **Y. Hodzhev**, G. Yordanov, R. Kalfin. Morphology and life cycle of the human blood microbiota in health and disease assessed by light and electron microscopy. Abstr. Book p. 254, M285, 10th FEMS Congress, 9-13 July 2023, Hamburg, Germany

15. Tsafarova, B., **Hodzhev Y.**, Tolchkov V., Youroukova V., Ivanova S., Kostadinov D., Yanev N., Panaiotov, S. Molecular analysis of sarcoid lung biopsys for *C. acnes* and *M. tuberculosis*. Poster N: 02405, 32 ESCMID Congress, Lisbon, Portugal.

Представяне на доклади и постери на национални конгреси

16. B. Tsafarova, **Y. Hodzhev**, V. Tolchkov, V. Youroucova, S. Ivanova, D. Kostadinov, N. Yanev, S. Panaiotov. PCR analysis for *Mycobacterium tuberculosis* complex shows negative results in sarcoid samples XV Национален конгрес на българските микробиолози с международно участие – Копривщица 05-08.10.2022.

17. **Й. Ходжев**, Б. Цафарова, В. Толчков, Н. Янев, С. Иванова, Д. Костадинов., В. Юрукова, С. Панайотов. Динамичен модел на микробиомна дисбиоза на кръвта при белодробна саркоидоза. XX Национален конгрес по клинична микробиология и инфекции на БАМ (16 - 18.09.2022), Пловдив.

18. **Й. Ходжев**, Б. Цафарова, В. Толчков, В. Юрукова, С. Иванова, Д. Костадинов, Н. Янев, Л. Григорова, С. Панайотов. Профилиране на кръвния микробиом при пациенти с белодробна саркоидоза: моделиране на микробиомната дисбиоза в момента на настъпване на заболяването. Осми конгрес на българското белодробно дружество по белодробни болести, 13-16.10.2022, Варна

19. **Й. Ходжев**, Б. Цафарова, В. Толчков, Н. Янев, В. Юрукова, С. Иванова, Д. Костадинов, С. Панайотов. Изследване на етиологията на саркоидозата: оценка на микробиома на дихателните пътища чрез метагеномен анализ на бронхоалвеоларен лаваж и биопсия на бронхиална тъкан. XXI конгрес на БАМ, 29 септември – 1 октомври 2023 г., гр. Пловдив

20. Б. Цафарова, **Й. Ходжев**, В. Толчков, В. Юрукова, С. Иванова, Д. Костадинов, Н. Янев, Л. Григорова, С. Панайотов. Методология за изолиране на ДНК и таргетно 16S и 18S

метагеномно секвениране на кръвни и биопсични проби от пациенти със саркоидоза. XX Национален конгрес по клинична микробиология и инфекции на БАМ – Пловдив 16-18.09.2022.

21. S. Panaiotov, B. Tsafarova, **Y. Hodzhev**, V. Tolchkov. Compounds causing the appearance of microbial-like structures in erythrocytes. Joint Forum: 11th South-East European Conference of Chemotherapy, Infections and Cancer and 31st Annual Assembly of International Medical Association Bulgaria. 28-31 October 2021, Medical University-Plovdiv, Bulgaria.

22. С. Панайотов, Б. Цафарова, **Й. Ходжев**, В. Толчков, Р. Калфин. Телцата на Хайнц имат предполагаем микробен произход. 19-ти Национален конгрес по клинична микробиология и инфектология на българската асоциация на микробиолозите. 14-16 септември 2021, София, България.

Участия като изпълнител в изследователски проекти:

1. „Разпространение и еволюция на ендемичен мултирезистентен генотип на *Mycobacterium tuberculosis* (SIT41, подгрупа TUR) в България, ФНИ - ДН13/4 от 15.12.2017, приключил март 2022 г., ръководител: проф. Стефан Панайотов

2. “Пещерите, като резервоар на нови и отново възникващи зоонози - екологично проследяване и метагеномен анализ в реално време” – договор ФНИ КП-06-ПН-51/8.

3. Национална научна програма „ВИХРЕН”, 2020 – 2025, проект „SARCOIDOSIS”, ФНИ - КП-06-ДВ/10 – ръководител: Проф. Стефан Панайотов

4. „Фундаментални, транслиращи и клинични изследвания в областта на инфекциите и инфекциозната имунология“- Европейския фонд за регионално развитие чрез Оперативна програма „Наука и образование за интелигентен растеж“ 2014-2020, договор BG05M2OP001-1.002-0001-C04

5. „Формиране и структурен анализ на еритроцитните грануларни телца“ – ФНИ КП-06-Н73/5 -5.12.2023 – Ръководител: проф. Стефан Панайотов

Обобщени наукометрични показатели

Научни публикации по темата на дисертацията	7
Научни публикации извън темата на дисертацията	8
Общо научни публикации:	15
Участия в международни научни форуми с постер или презентация по темата на дисертацията	15
Участия в национални научни форуми с постер или презентация по темата на дисертацията	7
Общо участия в научни форуми:	22

Общ брой кредитни точки за периода на докторантурата - 10.2020 - 10.2023: 1239 т.

Литература

1. Blaxter, M., Mann, J., Chapman, T., Thomas, F., Whitton, C., Floyd, R., & Abebe, E. (2005). Defining operational taxonomic units using DNA barcode data. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1462), 1935–1943. <https://doi.org/10.1098/rstb.2005.1725>
2. Breitwieser, F. P., & Salzberg, S. L. (2020). Pavian: Interactive analysis of metagenomics data for microbiome studies and pathogen identification. *Bioinformatics*, 36(4), 1303–1304. <https://doi.org/10.1093/bioinformatics/btz715>
3. Hodzhev, Y. (2023). Analysis of blood microbiome dysbiosis in pulmonary sarcoidosis by decision tree model. *Biotechnology & Biotechnological Equipment*, 37(1), 2283133. <https://doi.org/10.1080/13102818.2023.2283133>
4. Hodzhev, Y., Tsafarova, B., Tolchkov, V., Youroukova, V., Ivanova, S., Kostadinov, D., Yanev, N., Zhelyazkova, M., Tsonev, S., Kalfin, R., & Panaiotov, S. (2023). Visualization of the individual blood microbiome to study the etiology of sarcoidosis. *Computational and Structural Biotechnology Journal*, 22, 50–57. <https://doi.org/10.1016/j.csbj.2023.10.027>
5. Lu, J., Rincon, N., Wood, D. E., Breitwieser, F. P., Pockrandt, C., Langmead, B., Salzberg, S. L., & Steinegger, M. (2022). Metagenome analysis using the Kraken software suite. *Nature Protocols*, 17(12), 2815–2839. <https://doi.org/10.1038/s41596-022-00738-y>
6. Panaiotov, S., Filevski, G., Equestre, M., Nikolova, E., & Kalfin, R. (2018). Cultural Isolation and Characteristics of the Blood Microbiome of Healthy Individuals. *Advances in Microbiology*, 08(05), 406–421. <https://doi.org/10.4236/aim.2018.85027>
7. Panaiotov, S., Hodzhev, Y., Tsafarova, B., Tolchkov, V., & Kalfin, R. (2021). Culturable and Non-Culturable Blood Microbiota of Healthy Individuals. *Microorganisms*, 9(7), 1464. <https://doi.org/10.3390/microorganisms9071464>
8. Sneath, P. H. A., & Sokal, R. R. (1973). *Numerical taxonomy: The principles and practice of numerical classification*. W. H. Freeman.
9. The Galaxy Community, Afgan, E., Nekrutenko, A., Grüning, B. A., Blankenberg, D., Goecks, J., Schatz, M. C., Ostrovsky, A. E., Mahmoud, A., Lonie, A. J., Syme, A., Fouilloux, A., Bretaudeau, A., Nekrutenko, A., Kumar, A., Eschenlauer, A. C., DeSanto, A. D., Guerler, A., Serrano-Solano, B., ... Briggs, P. J. (2022). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. *Nucleic Acids Research*, 50(W1), W345–W351. <https://doi.org/10.1093/nar/gkac247>
10. Tsafarova, B. (2021). SARCOIDOSIS. *PROBLEMS of Infectious and Parasitic Diseases*, 49(1), 32–36. <https://doi.org/10.58395/pipd.v49i1.55>
11. Tsafarova, B., Hodzhev, Y., Yordanov, G., Tolchkov, V., Kalfin, R., & Panaiotov, S. (2023). Morphology of blood microbiota in healthy individuals assessed by light and electron microscopy. *Frontiers in Cellular and Infection Microbiology*, 12, 1091341. <https://doi.org/10.3389/fcimb.2022.1091341>

Microbiome and bioinformatic analysis for investigation of the sarcoidosis pathogenesis

Yordan Hodzhev

Abstract

Introduction

Pulmonary sarcoidosis, characterized by granulomas in lung tissue, poses significant therapeutic challenges due to its not fully uncovered etiology. The classical approach to investigating blood microbiota involves cultivation methods, which are designed to detect microbial cells in the blood that are latent but poses the potential for proliferation. Recent advancements in next-generation metagenomic sequencing (mNGS), coupled with bioinformatic tools, have enabled a deeper investigation of the blood microbiome, potentially elucidating the etiological factors of sarcoidosis.

Aim

Our study aims to delineate the association between blood microbiome dysbiosis and pulmonary sarcoidosis, employing a combination of cultivation methods, metagenomic sequencing, microbiome analysis, and advanced machine learning techniques.

Methods

For the purpose of addressing DNAemia, venous blood from 28 healthy adults was cultured, and metagenomes of cultured and non-cultured blood were compared. Culturing conditions included a vitamin K-rich medium and incubation at 43°C. Further, we analysed fresh blood samples from 7 patients with pulmonary sarcoidosis and 22 healthy individuals. Employing 16S rRNA gene amplicon sequencing, we characterized the microbiome in all described samples. Bioinformatic analyses, including taxonomic classification using Kraken 2 software, were complemented by machine learning algorithms. We developed a decision-tree supervised machine learning hierarchical model to discern patterns of microbial diversity and dysbiosis.

Results

The comparison between cultured and non-cultured blood from healthy individuals revealed significant differences, supporting the hypothesis that the isolated metagenomic DNA was from bacterial L-forms rather than just fragments of destroyed bacteria. The cultured samples showed a more diverse microbial profile compared to non-cultured samples, indicating the presence of viable bacteria in the bloodstream of healthy individuals. The bacterial and fungal compositions varied substantially between cultured and non-cultured samples, with dominant bacterial phyla including Proteobacteria, Firmicutes, Actinobacteria, and Planctomycetes in non-cultured samples, exhibiting a more diverse profile in cultured samples. Notably, certain bacterial genera such as *Veillonella*, *Prevotella*, *Cutibacterium*, *Corynebacterium*, and *Streptococcus* were either uniquely present or significantly augmented in patients. The machine learning model successfully identified and differentiated these patterns, highlighting its potential in diagnosing and understanding sarcoidosis underlying the phyla Proteobacteria, Firmicutes and Bacteroidetes and the genera *Veillonella* and *Comamonas* as pivotal taxa in classification of sarcoidosis.

Conclusions

This study emphasizes the critical role of the blood microbiome in understanding the etiology of pulmonary sarcoidosis. By integrating microbiome culturing, sequencing, and advanced machine learning analysis, we have unveiled significant insights into the disease's microbial underpinnings. These findings suggest a promising avenue for developing novel diagnostic and therapeutic strategies focused on microbiome modulation. However, further research with larger cohorts and diverse methodologies is essential to validate and expand upon these findings, ultimately enhancing our understanding of sarcoidosis and similar granulomatous lung diseases.

Keywords: pulmonary sarcoidosis, blood microbiome, machine learning, metagenomic sequencing, bioinformatic analysis