



НАЦИОНАЛЕН ЦЕНТЪР ПО ЗАРАЗНИ И ПАРАЗИТНИ БОЛЕСТИ

Деян Валентинов Дончев

**МЕТАГЕНОМНИ ПРОУЧВАНИЯ ВЪРХУ
ДЕТЕРМИНАНТИТЕ НА АНТИБИОТИЧНА
РЕЗИСТЕНТНОСТ В МОДЕЛНИ ПРОБИ ОТ
ОКОЛНА СРЕДА**

АВТОРЕФЕРАТ

на дисертационен труд

за присъждане на образователна и научна степен “Доктор”

Професионално направление: 4.3. „Биологически науки“

Докторска програма по „Микробиология“

Научен ръководител:

доц. Иван Иванов, дм

София, 2024 г.



НАЦИОНАЛЕН ЦЕНТЪР ПО ЗАРАЗНИ И ПАРАЗИТНИ БОЛЕСТИ

Деян Валентинов Дончев

**МЕТАГЕНОМНИ ПРОУЧВАНИЯ ВЪРХУ
ДЕТЕРМИНАНТИТЕ НА АНТИБИОТИЧНА
РЕЗИСТЕНТНОСТ В МОДЕЛНИ ПРОБИ ОТ
ОКОЛНА СРЕДА**

АВТОРЕФЕРАТ

на дисертационен труд

за присъждане на образователна и научна степен “Доктор”

Професионално направление: 4.3. „Биологически науки“

Научна специалност: 01.06.12. Микробиология

Научен ръководител:

доц. Иван Иванов, дм

София, 2024 г.

Дисертационният труд е разработен в Отдел “Микробиология”, в Национална Референтна Лаборатория по “Контрол и Мониториране на Антибиотичната Резистентност”, към НЦЗПБ, гр. София. Проучванията са финансирани от Националния фонд за научни изследвания на Република България чрез проект КП-06-Н23/5 и Европейския фонд за регионално развитие чрез Оперативна програма "Наука и образование за интелигентен растеж" 2014-2020; проект BG05M2OP001-1.002-0001-C04 "Фундаментални, транслиращи и клинични изследвания в областта на инфекциите и инфекциозната имунология", както и от асоциация Биокампус София.

Вътрешна защита се състоя пред разширен колегиум на НЦЗПБ на 10.10.2024 г. от 13:00 ч.

Дисертационният труд се състои от увод, цели и задачи, материали и методи, резултати и обсъждане, изводи, приноси, заключение, приложения и списък на цитираната литература и съдържа 203 страници, 30 фигури, 19 таблици и 2 приложения. Библиографията наброява 540 литературни източника, от които 7 на български език, 493 на английски език и 40 интернет страници.

Дисертационният труд е насрочен за публична защита на 21.03.2025 г. от ч. в аулата на НЦЗПБ, бул. Янко Сакъзов № 26, гр. София, на открито заседание на научното жури, назначено със заповед № 315/17.12.2024 г. на Директора на НЦЗПБ, съгласно ППЗРАСРБ на НЦЗПБ и ЗРАСРБ.

ИЗПОЛЗВАНИ СЪКРАЩЕНИЯ

16S-БД – база данни от гени 16S рРНК

АДИ – анализ на диференциалното изобилие

АМР – антимикробна резистентност

АСВ – ампликонов секвенционен вариант

БЛРС – бета-лактамаза с разширен спектър

БРА – бактерия, резистентна към антибиотици

ВФ – вакуум филтриране

ГАР – ген за антимикробна резистентност

КНИ – коефициент на надеждност на измерване

ЛПП – логаритмична промяна в пъти

МАГ – метагеном-асемблиран геном

МЛСБ – макролид, линкозамид, стрептограмин Б

МПК – минимална потискаща концентрация

ОТЕ – операционна таксономична единица

ПГЕ – подвижен генетичен елемент

ПСОВ – пречиствателна станция за отпадъчни води

ТОТ – точност на определяне на таксон

ФОМ – флокуляция с обезмаслено мляко

ХГП – хоризонтален генен пренос

ЧОТ – честота на откриване на таксон

СЪДЪРЖАНИЕ

1.	УВОД	6
1.	ЦЕЛИ И ЗАДАЧИ	8
3.	МАТЕРИАЛИ И МЕТОДИ	9
3.1.	Проби.....	9
3.1.1.	Проби за ампликоново метагеномно проучване	9
3.1.2.	Проби за цялостно метагеномно проучване.....	10
3.1.3.	Микробен тест стандарт	12
3.2.	Изолиране на ДНК.....	13
3.2.1.	In-house протокол за изолиране на ДНК.....	13
3.2.2.	Изолиране на ДНК чрез комерсиални протоколи	15
3.3.	Концентриране на биомаса.....	15
3.3.1.	Вакуум филтриране.....	15
3.3.2.	Флокуляция с обезмаслено мляко	16
3.4.	Секвениране.....	16
3.4.1.	Протокол за 16S секвенционни библиотеки	16
3.4.2.	Цялостно метагеномно секвениране	19
3.5.	Биоинформатични методи.....	20
3.5.1.	Качествен контрол на суровите прочити	20
3.5.2.	Оценяване на таксономичния състав	20
3.5.3.	Таксономична идентификация	21
3.5.4.	Анализи на алфа и бета разнообразие	22
3.5.5.	Анализ на диференциалното изобилие	22
3.5.6.	Асемблиране и ко-асемблиране.....	22
3.5.7.	Биниране.....	22
3.5.8.	Детекция на ГАР, ПГЕ и плазмиди	23
3.5.9.	Оценка на рисков фактор за човешко здраве.....	23
3.5.10.	Анализ на мрежите	23
4.	РЕЗУЛТАТИ И ОБСЪЖДАНЕ	24
4.1.	Оптимизиране на in-house протокол за изолиране на ДНК	24
4.2.	Дизайн за изпитване на ФОМ и ВФ.....	25

4.3. Резултати от сравнителни анализи върху in-house протокола, 16S-БД и ФОМ	25
4.3.1. КНИ стойности и отклонения от протоколите за изолиране на ДНК	26
4.3.2. Отклонения от 16S-БД и/или наборите за изолиране на ДНК върху проби МТС	26
4.3.3. Ефикасност на идентифициране на 16S базите данни при de-novo клъстериране.....	31
4.3.4. Оценка на флокулацията с обезмаслено мляко	32
4.4. Анализи, базирани на цялостни метагеномни проучвания	37
4.4.1. Сурови данни и МТС	37
4.4.2. Влиянието на ПСОВ Самоков върху р. Искър	37
4.4.3. Анализи, базирани на ПСОВ гр. Стара Загора и ПСОВ птицекланица „Евровърхът“	51
4.4.4. Анализи, базирани на сондажни (подпочвени) води в региона на гр. Стара Загора	62
5. ИЗВОДИ	68
6. ЗАКЛЮЧЕНИЕ.....	69
7. ПРИНОСИ	71
8. СПИСЪК НА ЦИТИРАНА ЛИТЕРАТУРА	72
9. СПИСЪК НА НАУЧНИ ТРУДОВЕ, СВЪРЗАНИ С ДИСЕРТАЦИОННИЯ ТРУД, И ПРОЕКТИ.....	74
10. БЛАГОДАРНОСТИ.....	75
11. ABSTRACT	76

1. УВОД

Антимикробната резистентност (АМР) представлява широко разпространено глобално предизвикателство [1]. АМР се появява, когато бактериите, вирусите, гъбичките и паразитите вече не се повлияват от антимикробни лекарства. В резултат на това антибиотиците и другите антимикробни лекарства стават неефективни и лечението на инфекциите се затруднява или става невъзможно, което увеличава риска от разпространение на болести, тежки заболявания, инвалидност и смърт. При различни инфекции, като инфекции на пикочните пътища, сепсис и инфекции, предавани по полов път, се отчитат високи и притеснителни нива на АМР в световен мащаб, което подчертава намаляващата ефективност на съществуващите препарати за лечение. През последното десетилетие мрежите за епидемиологично наблюдение в Европа отбелязват значителен скок в разпространението на бактерии, резистентни към антибиотици (БРА) [2]. Според статистически модели през 2019 г. са били регистрирани около 4.95 милиона смъртни случая, асоциирани с бактериална АМР, като 1.27 милиона от тях ще се дължат пряко на нея [1].

АМР и БРА са плод на дългогодишна неправилна и/или неконтролирана употреба на антимикробни лекарства в човешкото здравеопазване, ветеринарството, селското стопанство и растениевъдството, както и на изхвърлянето на тежки метали, което също допринася за развитието на АМР. Голяма част от антибиотиците се отделят от организма, непроменени в активна молекулна форма или като междинни метаболити, чрез фекалии или урина, животинска тор или друг вид биомаса по директен или индиректен път в околната среда или по-конкретно във водите, почвата и въздуха [3], където могат дори след многократно разреждане до суб-летални концентрации да селектират АМР. Отпадъчните води от големите градове и индустрията се събират посредством градски канализационни мрежи, които ги отвеждат към пречиствателни станции за отпадъчни води (ПСОВ), където те се пречистват механично, химично и биологично. След пречистването, рециклираната вода се зауства във водни басейни като реки, езера, потоци, морета и океани. В селското стопанство и животновъдството почвата е в директен контакт с отделената животинска тор, а антибиотиците се депозират върху почвата, от където впоследствие се просмукват надълбоко и попадат в подпочвени води. В резултат, околната среда около гъсто населите региони, такива с интензивно животновъдство и интензивна индустриална промишленост, са в пряк контакт с БРА, АМР и антимикробни вещества, които да способстват допълнително развитието на АМР. Дори, през последните години се възприе нормата, че околната среда е резервоар за АМР.

Не са напълно оценени всички рискове за общественото здраве, които АМР и БРА в околната среда крият. Разбира се, заразяването обратно на хора и животни с микроорганизми, развили АМР, е нисък, но вече редица проучвания показват, че такава опасност съществува посредством водни спортове, открито плуване в естествени водни басейни, пиене от естествени сладководни източници без пречистване, както и обработване на почвата в селското стопанство, напояване с

непречистени води в растениевъдството и вдишване на аерозоли. А редица други изследвания докладват разнообразието на БРА и гени за антибиотична резистентност (ГАР), както и разнообразие от подвижни генетични елементи (ПГЕ) в разнообразни локации като ПСОВ и сладководни басейни [4]. Европейските мрежи за надзор не остават безучастни и са задействали програми като EMBARK (Establishing a Monitoring Baseline for Antimicrobial Resistance in Key environments) и ревизираната Директива за пречистване на градските отпадъчни води [5]. Те стимулират внедряването на национално мониториране на големи обекти, които се явяват събирателни или горещи точки за развитие на AMP, като ПСОВ и други такива, обслужващи голям брой от населението.

България няма активно мониториране на AMP в околната среда и в бъдеще предстои да бъдат приети такива мерки. За тази цел най-важно е да бъде изграден капацитет, да се подготвят и обучат експерти, които да извършват този тип национално наблюдение. С решаването на тези проблеми са свързани и задачите на настоящия дисертационен труд.

Откриването на БРА и/или ГАР в конкретни точки може да се извърши по множество методики, всички с различни предимства и недостатъци. Сред тях метагеномните проучвания се открояват като най-модерни, най-информативни и с най-голямо поле за развитие и разширяване на техния информативен спектър, въпреки че са скъпоструващи и изискват трениран експертен персонал за анализиране на резултатите. В сравнение с културелните методи и тези, базирани на количествена полимеразна верижна реакция, тези методи предоставят в пъти по-богата информация. На територията на Република България, такъв тип проучвания с фокус AMP не са извършвани, което подчертава актуалността на труда.

Настоящото проучване си поставя за цел да бъде изследвано разпространението и вида на детерминантите на антимикробна резистентност, подвижните генетични елементи и бактериалното биоразнообразие в проби от околна среда и потенциалния риск за здравето на човека. Към групата потенциални проби от околна среда спадат ПСОВ, реки, езера, язовири и подпочвени води. Поставената цел е доста амбициозна и за нейното изпълнение бяха необходими първоначално редица проучвателни анализи. Тези първоначални изследвания, включващи множество оптимизационни и валидиращи стъпки, бяха адаптирани и имплементирани в протоколи за преданалитична подготовка на проби, както и в последващ биоинформатичен анализ. По този начин беше изграден капацитет за бъдещи проучвания в Националната референтна лаборатория по „Контрол и Мониториране на Антибиотичната Резистентност“ към Националния център по заразни и паразитни болести.

1. ЦЕЛИ И ЗАДАЧИ

Цел

Да се проучат разпространението и вида на детерминантите на антимикробна резистентност, подвижните генетични елементи и бактериалното биоразнообразие в проби от околна среда и потенциалния риск за здравето на човека.

Задачи

1. Създаване, оптимизиране и валидиране на цялостни алгоритми за преданалитична подготовка на проби за таргетно 16S рРНК (V3-V4 участъци) и цялостно (shotgun) секвениране при метагеномни проучвания като се оценят критичните стъпки.
2. Оптимизиране на собствен (*in-house*) метод за надеждно изолиране на геномна ДНК, подходяща за метагеномни проучвания и валидирането му чрез сравнителен анализ спрямо утвърдени протоколи.
3. Изпитване на метода *Флокулация с обезмаслено мляко* за концентриране на бактериална биомаса от водни проби чрез сравнителен анализ спрямо вакуум филтриране.
4. Извършване на сравнителен биоинформатичен анализ на съвременни 16S рРНК бази данни за идентификация при таргетно метагеномно проучване.
5. Извършване на дизайн на проучване чрез подбор на подходящи локации от околна среда с екологично значение, за които съществуват данни, че могат да съдържат гени за антимикробна резистентност и/или резистентни бактерии.
6. Провеждане на биоинформатичен анализ на данни, получени от проби от реки, подпочвени води и пречиствателни станции за определяне на таксономичното биоразнообразие, идентификация на потенциални патогени и анализ върху динамиката на биоразнообразието.
7. Анализ на гени за антимикробна резистентност, подвижни генетични елементи и оценка на потенциала на гените за антимикробна резистентност за хоризонтален генетичен пренос. Оценяване на динамиката, определяне на потенциален бактериален гостоприемник и риска за човешкото здраве.
8. Реконструирани и идентифицирани на метагеном-асемблирани геноми и депозиране на получените метагеномни данни в световните бази данни.

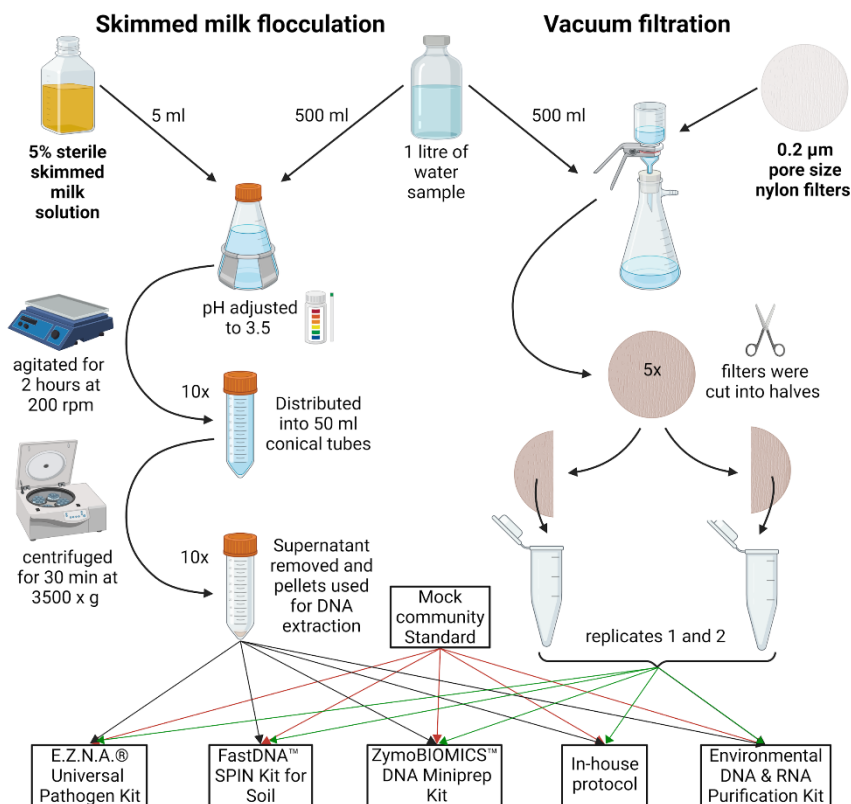
3. МАТЕРИАЛИ И МЕТОДИ

3.1. Проби

Пробите бяха събрани в стерилни HDPE пластмасови контейнери в различни обеми. Обемът варираше от 500 мл до 3 л при различните проби според тяхната чистота и техния произход, като от визуално по-чистите проби бяха събрани по-големи обеми. Времето и начинът на съхранение на пробите варираше, но в описаните граници.

3.1.1. Проби за ампликоново метагеномно проучване

На *Фигура №1* е представен дизайнът на ампликоновото метагеномно проучване.



Фигура №1. Схематично представяне на етапите на пречистване на вода чрез двете методики Флокулация с обезмаслено мляко и вакуум филтриране.

Накратко, беше взета 1 л некомпозитна проба вода от река Перловска, гр. София (координати 42.692164, 23.343892) и беше транспортирана в рамките на 30 мин до лабораторията. Тя беше разделена на две равни части от 500 мл за третиране с вакуум филтриране (ВФ) и флокулация с обезмаслено мляко (ФОМ). Тези две части бяха допълнително разделени на пет порции, всяка от които беше подложена в две повторения на различен протокол за изолиране на ДНК, в резултат на което бяха получени общо десет вида ДНК проби. Епруветките с утайки/филтри бяха ресуспендирани в лизиращия буфер от всеки набор за изолиране, както е показано на *Фигура №1*).

3.1.2. Проби за цялостно метагеномно проучване

Общо три метагеномни проучвания бяха изготвени, всеки с отделен дизайн и планиране. Всички проби, участващи в проучването са описани в *Таблица №1*.

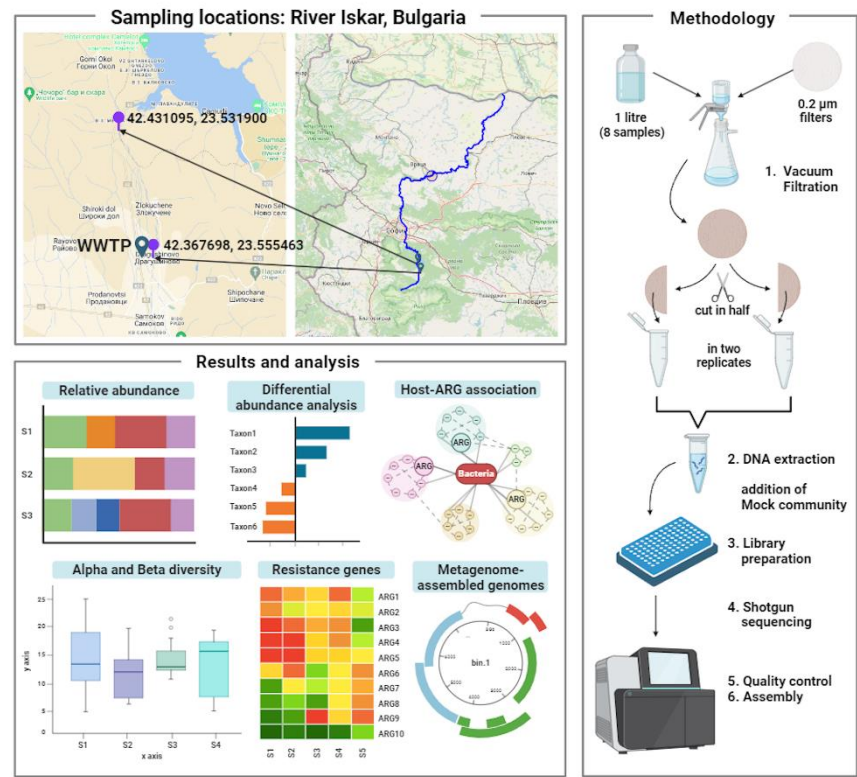
Таблица №1. Проби за цялостно метагеномно изследване.

№	Група	Цел	Вид проба	Локация	Дата
S1	река Искър	Динамика на ГАР, ПГЕ и влияние на ПСОВ Самоков върху р. Искър	повърхностна вода	с. Драгушиново	11/3/2022
S2			повърхностна вода	в.м. Мечката	11/3/2022
S3			повърхностна вода	с. Драгушиново	11/17/2022
S4			повърхностна вода	в.м. Мечката	11/17/2022
S5			повърхностна вода	с. Драгушиново	12/8/2022
S6			повърхностна вода	в.м. Мечката	12/8/2022
S7			повърхностна вода	с. Драгушиново	12/21/2022
S8			повърхностна вода	в.м. Мечката	12/21/2022
MTC			синтетичен метагеном		
S1	Подпочвени води около гр. Стара Загора	Наличие на ГАР, ПГЕ, плазмиди в подпочвени води	сондажна вода	аквапоника	4/17/2024
S2			утайтелна шахта	аквапоника	4/17/2024
S5			сондажна вода	кравеферма	4/17/2024
S6			сондажна вода	свинеферма	4/18/2024
S12			сондажна вода	овцеферма	4/17/2024
S8	ПСОВ, Стара Загора	ГАР, ПГЕ, плазмиди, влизаци и излизаци от градска ПСОВ	входяща, мръсна	ПСОВ гр. Стара Загора	4/17/2024
S10			активна утайка	ПСОВ гр. Стара Загора	4/17/2024
S7			изходяща, рециклирана	ПСОВ гр. Стара Загора	4/17/2024

S14	ПЦОВ,	ГАР, ПГЕ, плазмиди, влизаци	входяща, мръсна	ПЦОВ, Птицекланица "Евровърхът"	4/25/2024
S15	Птицекланица	и излизаци от локална ПЦОВ	изходяща, рециклирана	ПЦОВ, Птицекланица "Евровърхът"	4/25/2024
MTC			синтетичен метагеном		

3.1.2.1. Река Искър след ПЦОВ гр. Самоков

Общо осем композитни повърхностни водни проби (по 1 л всяка) бяха взети по двойки по поречието на река Искър от двете места (42.367698, 23.555463 - с. Драгушиново и 42.431095, 23.531900 - вилна зона "Мечката") с автоматичен пробовземач Bühler 2000 Portable (Hach UK, Manchester, UK) за период от 24 ч, за да се избегне отклонението ден-нощ на следните четири дати: 3 ноември 2022 г., 17 ноември 2022 г., 8 декември 2022 г. и 22 декември 2022 г. В резултат на това бяха взети общо 8 проби (2 места × 4 дати).



Фигура №2. Графичен абстракт на shotgun метагеномното проучване върху влиянието на отпадъчните води от ПЦОВ Самоков върху река Искър.

Първото място се намира на около 800 м след градската ПСОВ на гр. Самоков, като по този начин е подходящо за мониториране на рециклираните води непосредствено след вливането им в реката, а второто място е на около 10 км надолу по течението, точно преди вливането на реката в язовир Искър (Виж *Фигура №2*). Пробите бяха транспортирани до лабораторията в рамките на 6 ч и обработени веднага. Първите четири проби от първите две дати бяха разделени на две порции от по 500 мл и обработени съответно по двете методики за концентриране на биомаса ВФ и ФОМ (Виж *Фигура №1*). Филтрите бяха съхранявани при -80°C, докато всички проби бяха налични.

3.1.2.2. Сондажни води в района на гр. Стара Загора

Сондажните води бяха събрани от най-близко достъпната чешма или крайник като предварително водата беше оставена 5 мин да се изтече. От всички източници на вода в групата „Подпочвени води“ от *Таблица №1* бяха събрани по 3 л в стерилни контейнери и транспортирани на 4°C в рамките на 4 ч в лабораторията и филтрирани по методика на ВФ. По-голям обем беше предвиден поради факта, че за разлика от останалите видове проби сондажната вода е много бистра и видимо няма частици, които да задръстват филтрите. На всички четири проби от подземна вода трите литра преминаха през филтрите без да бъде задръстен. Към тази група проби на подпочвена вода беше добавена и проба S2, която е събрана от изходната точка на обекта аквапоника от утайтелната шахта, накрая на пътя на водата. В резултат, от обект аквапоника може да се наблюдава състава на входящата чиста вода (S1) и на изходящата мръсна вода (S2). Точните локации са както следва: S1 (аквапоника, сондаж) и S2 (аквапоника, утайтелна шахта) - 42.403480, 25.571982; S5 (кравеферма, сондаж) - 42.409400, 25.646555; S6 (свинеферма, сондаж) - 42.411781, 25.649965; S12 (овцеферма, сондаж) - 42.349836, 25.593113.

3.1.2.3. ПСОВ гр. Стара Загора и ПСОВ птицекланица

С разрешението и помощта на екипа на ПСОВ гр. Стара Загора (42.408104, 25.675919) бяха събрани проби от трите ключови точки: входяща мръсна вода (S8), активна утайка (S10) и рециклирана чиста вода (S7). Рециклираната чиста вода се зауства в р. Бидечка.

Птицекланица „Евровърхът“ (42.414614, 25.660605) разполага със собствена малка ПСОВ за целите на клането на птици. С разрешението на управата бяха събрани две проби – една от входящата мръсна вода (S14) и изходяща чиста (S15), както е посочено на *Таблица №1*.

3.1.3. Микробен тест стандарт

Микробен тест стандарт (MTC, от англ. “Mock community standard” или “Microbial community standard”) е клетъчна суспензия от набор от неживи, но интактни и стабилизирани клетки на микроорганизми, обикновено референтни щамове, които са обединени в изкуствено микробно съобщество в строго определени концентрации помежду си. В случая беше използван ZymoBIOMICS Microbial

Community Standard (кат. № D6300, Zymo Research Corporation, USA) с 8 бактерии и 2 дрожди. Този стандарт представлява положителна контрола, която се третира като проба и преминава всички стъпки от изолиране на ДНК до секвениране, по която след биоинформатичния анализ се съди и оценява степента на отклоненията, които се въвеждат с тези стъпки. Същият МТС беше използван за сравнителния анализ на ДНК протоколите за изолиране и за сравнителния анализ на базите данни за 16S рРНК идентификация. Също така, при всички цялостни метагеномни проучвания, МТС беше успоредно третиран като реална проба като контрола на екстракцията.

3.2. Изолиране на ДНК

3.2.1. *In-house* протокол за изолиране на ДНК

In-house протоколът беше разработен чрез комбиниране на елементи от други публикувани протоколи [6,7] с леки модификации и беше оптимизиран в лабораторията. На Таблица №2 са представени необходимите консумативи, а на Таблица №3 – стъпките на протокола. Протоколът е продукт от множество оптимизационни етапа и описаният метод е крайният продукт.

Таблица №2. Необходими реактиви и консумативи преди началото на *in-house* протокола за изолиране на ДНК.

Буфери и реагенти, необходими преди протокола:	Бележки
ML-1 буфер , съдържащ: 0.120M GuSCN (гуанидин изотиоцианат) и 0.180M Na ₂ HPO ₄ (динатриев фосфат)	Филтриран през 0.22 µm филтър. Може да се приготви предварително.
ML-2 буфер , съдържащ: 0.15M NaCl (натриев хлорид), 0.2 % SDS (натриев додецилсулфат), 0.455 M TRIS-HCl и 0.0455 M TRIS-база	Филтриран през 0.22 µm филтър. Може да се приготви предварително.
ML-4 буфер , съдържащ 0.120M (NH ₄)Al(SO ₄) ₂ (амониев алуминиев сулфат додекахидрат)	Филтриран през 0.22 µm филтър. Може да се подготви предварително.
Циркониеви перли – 0.1 мм и 0.5 мм	В съотношение 1:1 от двата размера в краен обем 250-300 µl в 2 мл епруветка за изолиране на ДНК.
Топчета от неръждаема стомана - 4 мм или 5 мм	По 1 топче на реакция. Добавени бяха в епруветката за изолиране на ДНК.
Пречистващ разтвор от 3-кромол-1-хлоропропан или хлороформ и изоамилов алкохол в съотношение 24:1	Необходимият обем за една проба е 500 µl.
PVSA (поливинилсулфонова киселина)	45 мг/мл в 1000x разтвор. Смесен предварително с ML-1 непосредствено преди изолирането.
5M NaCl	Затоплен на гореща баня, за да се разтвори в случай на кристализация.
1% Tween-20	

Магнитни частици SeraMag	Темпераиран на стайна температура за 30 мин преди употреба.
SiO2 грес (грес/смазочен материал за висок вакуум),	Автоклавиран на малки партии при 121 °C, 118 psi, 15 мин. Прехвърлен в спринцовка. Греста разделя фазите по време на разделянето на белтък-ДНК, което позволява лесно преливане на супернатантата, съдържаща ДНК, без да се докосва белтъчната фаза.
99 % изопропанол	Съхраняван на 4°C докато е необходим.
70% етанол за промиване.	Необходими са 1.5 мл за всяка проба. Съхраняван при 4°C докато е необходим
Епруветки за изолиране на ДНК от 2 мл	ВАЖНО! Използвани са висококачествени центрофужни епруветки. Дъното на епруетката може да се счупи от стоманеното топче и скоростта на центрофугата.

Таблица №3. Стъпки на *in-house* протокола.

	Стъпки
1	Подготвени бяха епруветки за изолиране на ДНК от 2 мл с 0.15 мм и 0.5 мм циркониеви перли в общ обем 300 µл в съотношение 1:1 и 1 стоманено топче.
2	Смесени бяха предварително ML-1 и PVSA с 360 µл ML-1, 0.5 µл PVSA на проба и 114 µл ML-4.
3	Филтрите бяха асептично нарязани на по-малки парчета и поставени в епруетките за изолиране на ДНК.
4	Добавени бяха 475 µл от сместа ML-1-PVSA-ML-4 към пробата и сместа беше хомогенизирана.
5	Добавени бяха 36 µл ML-2 и сместа беше хомогенизирана.
6	Епруетките бяха плътно затворени и хомогенизирани на BeadBeater за 2x1 мин при 25 об./мин с пауза между всяка минута, за да се охладят епруетките. Тази стъпка е беше допълнително оптимизирана за постигане на най-високо качество на ДНК по отношение на добива и крайния размер.
7	Епруетките бяха центрофугирани за кратко, за намаляване на пияната и отстраняване на перлите от капачката.
8	Добавени бяха 18 µл от 5M NaCl и 500 µл от разтвора за пречистване.
9	Добавени бяха SiO2 грес с размер на грахово зърно във вътрешната част на капачката. Сместа беше хомогенизирана.
10	Пробите бяха поставени вертикално върху поставка за лед за 10 мин. Извадени бяха магнитни частици SeraMag от хладилника за темперирание.
11	Центрофугирани бяха пробите при 15 000 об./мин за 15 мин.
12	Изчислен беше необходимия първоначален обем на SeraMag частиците с 8 µл SeraMag топчета на проба + 1. Необходимият обем беше прехвърлен в чиста епруетка от 1.5 µл и поставен върху магнит за разделяне. Супернатантата беше изхвърлена.
13	Бяха извършени две стъпки на промиване с поне 3 пъти по-голям обем от 1% Tween-20, като междувременно са хомогенизирани чрез бавно пипетиране и поставени върху магнита. Не се изисква инкубиране между промивките.
14	Хомогенизиране на окончателно промитите магнитни частици в 2x първоначалния обем на 1% Tween-20 (8 µл в 16 µл).
15	В нов набор епруетки бяха разпределени по 16 µл магнитни частици в Tween-20.

16	Супернатантата (около 480 μ л) от епруетките за изолиране на ДНК от центрофугата беше пипетирана директно върху прясно ресуспендираните магнитни частици. Обемът на супернатантата може да варира, затова беше отбелязан обема й.
17	0,7x обема охладен изопропанол бяха добавени към всяка супернатанта (336 μ л изопропанол за 480 μ л SN) и сместа беше хомогенизирана.
18	Пробите бяха инкубирани за поне 10 мин на стайна температура.
19	По време на инкубацията беше приготвен разтвор за промиване със 70% етанол от 1.5 мл за всяка проба.
20	Поставени бяха епруетките върху магнит, докато се избистрят.
21	Супернатантата беше изхвърлена.
22	Извършени бяха два промивки на магнитните частици със 750 μ л 70 % етанол с инкубация за минимум 30 с. Без хомогенизиране и без премахване от магнита по време на промиването.
23	Последният обем етанол беше отстранен с бавно пипетиране, за да се избегне оставянето на капки по стените на епруетката. Остатъчният етанол от дъното на епруетката също беше отстранен.
24	Магнитните частици бяха оставени на въздух да се изпари остатъчният етанол за около 1:30 $\div$ 2 мин, като епруетките бяха отворени. Магнитните частици не бяха пресушени, тъй като това застрашава възстановяването на ДНК.
25	Частиците бяха разтворени в 50 $\div$ 100 μ л 10mM Tris/NF вода.

3.2.2. Изолиране на ДНК чрез комерсиални протоколи

Бяха подбрани общо 4 набора за изолиране на ДНК по фактори като най-добри резултатите при други сравнителни изследвания, цена и достъпност. Това бяха ZymoBiomix DNA miniprep kit (Zymo Research Corporation, USA), EZNA Universal Pathogen kit (Omega Bio-tek, Inc. USA), FastDNA SPIN kit for Soil (MP Biomedicals, UK), Environmental DNA & RNA Purification kit (EURX Sp. z o.o.). Всички протоколи бяха извършени съгласно инструкциите на производителя с изключение, че накрая геномната ДНК беше разтворена във вода, чиста от РНКази, вместо в елуиращите буфери, предоставени в китовете. Заедно с новоразработения *in-house* метод тези пет протокола за изолиране на ДНК бяха приложени върху пробите за ампликоново секвениране както и пробите МТС. ДНК от пробите, подложени на цялостно метагеномно секвениране, беше изолирана чрез Environmental DNA & RNA Purification kit (EURX Sp. z o.o.).

ДНК на всички проби беше измерена както спектрофотометрично с BioDrop μ Lite+ (Biochrom Ltd.), така и флуорометрично с Qubit 4.0 Fluorometer (Thermo Fisher Scientific, САЩ).

3.3. Концентриране на биомаса

3.3.1. Вакуум филтриране

Пробите бяха подложени на ВФ през найлонови филтри с диаметър 47 мм и размер на порите 0.2 μ м (Cytiva, WhatmanTM), като се използва филтрираща система Lafil 400 - LF 30 (Rocker). Всеки филтър беше стерилно разрязан на две равни половини, за да бъде допълнително екстрахиран в две повторения. Всяка половина беше допълнително нарязана на по-малки парчета за по-добро лизиране и допир от

перлите в епруветките за изолиране. Нарязаните филтри бяха директно добавени в епруветки за изолиране на ДНК.

3.3.2. Флокуляция с обезмаслено мляко

ФОМ беше извършен по предварително описан протокол [8]. Накратко, сухо обезмаслено мляко на прах (HiMedia Laboratories, India) беше разтворено в 5% разтвор и автоклавирано за 15 мин при 115°C, 18 psi. От него 5 мл бяха добавени към 500 мл проба, за да се постигне 0.05% крайна концентрация. РН на пробата беше коригирана до 3.5-4.0 с 1M HCl, поставена на хоризонтална клатачна машина и разбъркана на 200 оборота в минута в продължение на 2 ч при стайна температура. След това беше разпределена в конични епруветки от 50 мл и центрофугирана при 3500 x g за 30 мин при 4°C. Супернатантата беше декантирана и епруветките бяха обърнати надолу, за да се изтече остатъчната вода за 5 мин. Утайките бяха подложиха на изолиране на ДНК.

3.4. Секвениране

3.4.1. Протокол за 16S секвенционни библиотеки

По литературни данни бяха подбрани праймерни двойки за вариабилните региони V3-V4 на гена 16S рРНК [9]. Праймерните двойки (Pro_341F 5'-CCTACGGGNNBGCASCAG-3' и Pro_805R_Japan 5'-GACTACNVGGGTATCTAATCC-3') бяха поръчани за синтез директно с нуклеотидни опашки, предвидени за платформата Illumina, към които чрез допълнителна амплификация се добавят индексни секвенции. Праймерните двойки бяха сравнени чрез *in-silico* анализ с други вече налични праймерни двойки като такива, включени в комерсиални протоколи за 16S библиотеки или тези от проекта „Земният микробиом“. На *Таблица №4* и *Таблица №5* е описан протоколът за 16S амплификация. На *Таблица №6* и *Таблица №7* са описани реактивите и стъпките за пречистването на получените ампликони. След това, беше изпълнена индексна ПВР (Виж *Таблица №8* и *Таблица №9*). Оптималното количество ампликонна ДНК, което беше добавено в индексната ПВР, беше между 15 и 60 нг. Пипетирането беше извършено с обеми от ≥ 2 μ л. Накрая получените индексирани ампликони бяха пречиствени чрез двойна селекция по размер и от реактивите от ПВР (*Виж Таблица №10*). Секвенционните библиотеки на всички проби бяха обединени в еквимоларни концентрации в общ пул, който беше секвениран на инструмент MiSeq V3 (2 x 300 нб).

Освен това за ДНК пробата на МТС, изолирана кита Zymo miniprep DNA, извършихме 16S амплификацията в три допълнителни реакции с градиентно нарастващи температури на хибридизация на праймерите (55°C, 58.5°C, 62°C), като 62°C-г2 се явяваше повторение на оригиналната в протокола. Така бяха получени 8 секвенционни библиотеки с МТС, едната от които е повторение на 62°C.

Таблица №4. Стъпки за подготовка на 16S рРНК амплификационна ПВР.

1	Приготвена беше една епруветка за мастрмикс от всички реактиви с изключение на DTT, праймерите и ДНК пробата. Дназата беше добавена последна. Сместа беше хомогенизирана внимателно чрез пипетиране.
2	Мастърмикс реакцията беше инкубирана при 37°C за 30 мин, за да може ензима ДНКаза да отстрани остатъчната ДНК.
3	Бяха добавени дитиотритол (DTT) и праймерите.
4	Мастърмикс реакцията беше инкубирана при 37°C за 30 мин, за да деактивирате ензима ДНаза.
5	Мастърмикс сместа беше разпределена по 18 µl в епруветки за ПВР.
6	ДНК от всяка проба беше добавена и хомогенизирана клатачна машина с помощта на приставка за ПВР плаки или чрез пипетиране. Плаката беше запечатана с пластмасово фолио.
7	Реакцията премина при следния протокол: 98°C за 10 мин (активиране на полимеразата); 25 цикъла при 96°C за 10 с (денатурация), 62°C за 30 с (хибридизация и синтез); 68°C за 5 мин за окончателно удължаване.

Таблица №5. Протокол за амплификация на гена 16S рРНК– амплификационна ПВР.

бр. реакции	10			
обем на реакцията	20			
Реагент	Начална конц.	крайна конц.	Обем за 1 реакция, [µl]	Краен обем за всички реакции [µl]
HYBRID buf [x] (Hybrid DNA Polymerase, EURx)	10	1	2.00	20.00
dNTP [mM] (Hybrid DNA Polymerase, EURx)	10	0.3	0.60	6.00
16S_Pro_NGS_primers [uM]	10	0.2	0.40	4.00
onHYBRID [U/ul] (Hybrid DNA Polymerase, EURx)	2	0.06	0.60	6.00
HNB.Na2 [mM] (Sigma)	5	0.055	0.22	2.20
DTT [mM] (EURx)	10	0.25	0.50	5.00
DNase NEB [U/ul] (New England Biolabs)	2	0.018	0.18	1.80
TPRA-Cl [mM] (Sigma)	300	32	2.13	21.33
DMSO [%] (Sigma)	100	2	0.40	4.00
MgCl2 [mM] (Hybrid DNA Polymerase, EURx)	25	0.5	0.40	4.00
sample DNA/NTC			2.00	20.00
mol. grade H2O (EURx)			10.57	105.67

Таблица №6. Необходими реагенти за стъпките на пречистване на ампликоните.

Буфери и реагенти	Бележки:
SDB буфер, съдържащ: 0.495% Ficoll 400; 0.0275% Dextran 500; и 5% Tween20	Обем на реакцията беше 20 µl. Буферът беше оставен на стайна температура за поне 20 мин преди употреба.

Магнитни частици MagBio (MagBio Genomics Inc.)	Темперирани на стайна температура за поне 20 мин преди употреба.
Буфер за ресуспендиране (RSB), от който и да е комплект за подготовка на библиотеки на Illumina или търговски комплект за подготовка на библиотеки или еквивалентен буфер като 10 mM Tris-HCl, pH 8.5 с 0.1% Tween 20	Предварително беше загрят до 70°C.
70% етанол за промиване	Необходими са 2 x 180 µл. Приготвен беше пресен в деня на анализа. Съхраняван на 4°C до момента на употреба

Таблица №7. Стъпки за изпълнение на единична селекция по размер.

1	Индексния ПВР продукт беше разреден с SDB буфер в съотношение 1:1 и хомогенизиран на шейкър/вортекс.
2	Добавени бяха 0.55x магнитни частици и хомогенизирани на шейкър/вортекс.
3	Сместа беше инкубирана в продължение на 5 мин при стайна температура.
4	Плаката беше поставена върху магнит за плаки и инкубирана докато суспензията стане прозрачна.
5	Супернатантата беше отстранена и изхвърлена.
6	Две последователни промивания бяха извършени на магнитните частици със 180 мл 70% етанол. Инкубирани бяха за минимум 30 с. Всеки път етанолът беше изхвърлян. Не бяха ресуспендирани частите в етанола и не беше отстранявана плаката от магнита по време на промиванията.
7	Последният обем етанол беше отстранен с бавно пипетиране, за да се избегне оставянето на капки по стените на епруветката. Остатъчният етанол на дъното на ямките след последното промиване също беше отстранен с пипетиране с малък обем.
8	Магнитните частици бяха подсушени на въздух за приблизително 90 ÷ 120 с, като плаката беше оставена отворена. Частиците не бяха пресушавани, тъй като това намалява възстановяването на ДНК.
9	Елуирани бяха в 25 µл предварително загрят RSB. Почистеният продукт може да се съхранява при 4°C за една нощ.

Таблица №8. Стъпки за изпълнение на индексна ПВР.

1	Беше подготвена една епруетка за мастърмикс от всички реагенти с изключение на ДНК на 16S ампликона и индексните праймери. Ресуспендирани бяха внимателно с пипета и разпределени по 13 µл от мастърмикс за всяка реакция в нова ПВР плака.
2	С пипета бяха нанасени 4 µл от UD индексите в ПВР плаката с мастърмикс.
3	С пипета бяха добавени по 3 µл 16S ампликонна ДНК от 16S амплификационен ПВР.
4	Плаката беше леко хомогенизирана на вортекс с помощта на приставка за плаки или чрез пипетиране. Плаката беше запечатана с пластмасово фолио.
5	ПВР апаратът беше настроен по следния протокол:
6	98°C за 10 мин (активиране на полимеразата); 8 цикъла при 95°C за 15 с (денатурация), 65°C за 25 с (хибридизация и синтез) и 70°C за 25 с (удължаване); 72°C за 5 мин за окончателно удължаване
7	Индексният ПВР продукт може да се съхранява при 4°C за една нощ.

Таблица №9. Протокол за индексна ПБР.

бр. реакции	10			
обем на реакцията	20			
Реагент	начална конц.	крайна конц.	Обем за 1 реакция [μl]	Краен обем за всички реакции [μl]
HYB buffer [x]	10	1	2.00	20.00
dNTP [mM]	10	0.2	0.40	4.00
UD_unique_index (Illumina DNA/RNA UD Indexes Set A)	5	1	4,00	40,00
MgCl2	25	0.5	0.40	4.00
onHYBRID_pol	2	0.06	0.60	6.00
16S amplicon (optimal - 15-60 ng)			3.00	30.00
mol. grade H2O			9.60	96.00

Таблица №10. Стъпки за извършване на двойната селекция по размер.

1	Обемът на индексната ПБР беше разреждана със SDB буфер в съотношение 1:1 и хомогенизирана на клатачна машина/вортекс.
2	Добавени бяха 0.40x обема магнитни частици и хомогенизирани отново.
3	Продуктът беше инкубиран в продължение на 5 минути при стайна температура.
4	ПБР плаката беше поставена на магнит до избистряне.
5	По време на инкубацията бяха разпределени по 0.50x обема магнитни частици в нова чиста плака.
6	Докато първата плака беше върху магнита, цялата супернатанта беше пипетирана и прехвърлена в новата плака с 0.50x обема магнитни частици. Пробите бяха хомогенизирани внимателно чрез пипетиране.
7	Продуктът беше инкубиран в продължение на 5 мин при стайна температура.
8	ПБР плаката беше поставена на магнит до избистряне.
9	След инкубацията супернатантата беше пипетирана и изхвърлена.
10	Бяха извършени две последователни промивания на магнитните частици със 180 мл 70% етанол с кратка инкубация за 30 с без хомогенизиране и отделяне от магнита.
11	Последният обем етанол беше отстранен с внимателно пипетиране, за да се избегне образуването на капки по стените на епруветката. Остатъчният етанол от дъното на епруветката след последното промиване беше допълнително отстранен.
12	Магнитните частици бяха подсушени на въздух за приблизително 90 ÷ 120 с, като се оставя отворена плаката. Частиците не бяха пресушавани, за да не се намали добива на ДНК.
13	ДНК беше разтворена в 25 μl предварително загрят буфер RSB.

3.4.2. Цялостно метагеномно секвениране

От всяка двойка повторния пробата с най-високо съотношение 260/280 за чистота и най-висока концентрация на ДНК бяха използвани за изграждане на НГС библиотеки с Colibri™ ES DNA Library Prep Kit (Thermo Fisher Scientific, САЩ) или Illumina DNA Prep kit (Illumina, Inc., San Diego, California, U.S.A.) с тотално

количество входяща геномна ДНК 50 нг. Библиотеките бяха секвенирани на NextSeq 550 с комплект V2.5 Mid/High Output (2 x 150 нб) (Illumina, Сан Диего, Калифорния).

3.5. Биоинформатични методи

Биоинформатичният анализ може да съвпада при двата вида метагеномни проучвания, но те имат различни аспекти. Поради това, различията са допълнително описвани при необходимост. При пробите за цялостно метагеномните секвениране различията произлизаха от различни версии на базите данни за ГАР и таксономична идентификация, защото са обработени по различно време.

3.5.1. Качествен контрол на суровите прочити

Необработените нуклеотидни прочити на всички проби бяха автоматично демултиплексирани и адаптерите за секвениране бяха премахнати в BaseSpace (Illumina) и от тук нататък наречени за улеснение сурови прочити.

3.5.1.1. Проби от ампликонното метагеномно секвениране

За обработване на суровите прочити на пробите от р. Перловска и МТС беше използвана програмата cutadapt v4.6, за да се премахнат праймерите на 16S rPHK и да се отстранят базите с ниско качество (<20 Q в 5'края и <15 Q в 3'края). В зависимост от подхода за клъстериране химерните прочити бяха отстранявани чрез приложението USEARCH за *closed-reference* OTE или vsearch при *de novo* клъстериране или автоматично при Ампликонов секвенционен вариант (ACB) от приложението DADA2.

3.5.1.2. Проби от цялостно метагеномно секвениране

Суровите прочити бяха автоматично подрязани въз основа на съдържанието на адаптери и средния резултат за качество Q30+ чрез приложението Trim-galore v0.6.7 с настройки по подразбиране, което гарантира, че се запазват само висококачествени прочити. След това бяха почистени от човешки прочити посредством приложението bmtagger чрез налагане върху човешкия геном и тези, който съответстваха, бяха отстранени. Останалите прочити бяха проверени за качество с FASTQC v0.11.9.

3.5.2. Оценяване на таксономичния състав

За оценяване на коефициент на надеждност на измерването (КНИ, от англ. “Measurement Integrity Quotient”) бяха използвани приложенията miqScore16SPublic [10], и miqScoreShotgunPublic [11] за съответно ампликоново и цялостно метагеномни проучвания. Те се прилагат върху МТС пробите. Те работят директно със суровите данни директно от секвенатора. Успоредно беше разработен и използван наш калкулатор в Python скрипт [12], който използва заложената формула за калкулиране на КНИ в оригиналните приложения. С този алтернативен калкулатор бяха оценени пробите МТС, понеже позволява анализ на таксономични таблици, генерирани по алтернативни подходи и не изисква суровите нуклеотидни прочити. За получаване на таксономична таблица приложението miqScore16SPublic генерира ACB групи чрез

DADA2, докато miqScoreShotgunPublic налага суровите прочити върху геномите на 10-те микроорганизма. След това формулата за КНИ се прилага върху таксономичните таблици.

КНИ е стойност, която се присъжда на дадена МТС проба, и варира между 0 (максимално отклонение) и 100 (липса на отклонение) въз основа на сравнението на наблюдавания и очаквания състав като резултати >90 се считат за отлични, а $80 \div 89$ - добри. Тя е обща калкулация от сравнението на получената честота на всеки микроорганизъм спрямо неговата теоретично очаквана.

3.5.3. Таксономична идентификация

3.5.3.1. При ампликоново метагеномно секвениране

Суровите прочити на пробите от р. Перловска и пробите МТС бяха анализирани в платформата Qiime2 посредством приложението miqScore16SPublic, както и нашия алтернативен калкулатор за КНИ за определяне на най-добрия кит за изолиране на ДНК, който предоставя най-близък до очаквания микробен състав. Резултатите от приложението miqScore16SPublic за изчисляване на КНИ бяха счетени за референтни.

Успоредно с това, суровите прочити бяха обработени с шест база данни за идентификация на гена 16S rRNA (16S-БД). Следните пълноразмерни БД бяха изтеглени Silva 99%-OTU (Silva), GSR-99%-OTU (GSR), GTDB-214.1-99%-OTU-less (GTDB-less), GTDB-214.1-99%-OTU-full (GTDB-full), GreenGenes-13_8-99%-OTU (GG_13.8), EzBio-Cloud-v.2018 (Ezbio). Референтните им последователности бяха извлечени въз основа на *in-silico* ПБП с Pro двойките праймери и използвани като входни данни за алгоритъма Naïve-bayes в Python пакета sk-learn v1.5.0 за обучение на персонализирани класификатори като бяха използвани параметри по подразбиране. Класификаторите бяха прилагани за идентификация при индекс на сигурност (от англ. „confidence score“) от 0.7 (по подразбиране).

Беше ползван подходът за клъстериране на операционна таксономична единица (ОТЕ). За целите на тестването бяха получени 99% ОТЕ клъстериранни проби (за всяка 16S-БД) по двата подхода *de novo* и *closed-reference*. *De novo* клъстерираните ОТЕ групи бяха обработени с приставката „evaluate-composition“ като се сравниха двата показателя: честота на откриване на таксони (ЧОТ, от англ. “taxon detection rate”) и точността на определяне на таксоните (ТОТ, от англ. “taxon accuracy rate”). От друга страна, клъстерираните чрез *closed-reference* ОТЕ проби бяха използвани за анализ на отклоненията, които се внасят в таксономичното разпределение от самите БД, като беше изчислен КНИ чрез нашия алтернативен калкулатор.

3.5.3.2. При цялостно метагеномно секвениране

Таксономичното определяне на суровите прочити беше извършено с Kraken2 v2.1.2 с база данни PlusPF (Standard plus Refeq Protozoa & fungi), създадена през януари 2023 г., последван от Bracken v2.8, както е описано в протокол [13]. Различни

параметри на филтриране, както и подходи за нормализиране бяха прилагани в зависимост от анализите. Те са описани под всяка фигура в секция Резултати и Обсъждане. Броят беше трансформиран в относителни честоти, а таксоните под 1.5% бяха обединени в "Други" и използвани за визуализиране на таксономичния състав с Python пакетите matplotlib v3.7 и seaborn v0.12.0.

3.5.4. Анализи на алфа и бета разнообразие

Анализите на алфа и бета разнообразието, груповата значимост и променливост бяха извършени с ResistoXplorer и Qiime2. Статистическите изчисления бяха извършени автоматично от същите приложения и платформи.

3.5.5. Анализ на диференциалното изобилие

Анализ на диференциалното изобилие (АДИ, от англ. „differential abundance analysis“) беше използван и при двата вида метагеномно секвениране. АДИ беше извършен с помощта на метода и приложението ANCOM-BC2 [14], което е интегрирано като приставка в платформата Qiime2. Понеже таксономичните таблици от пробите за цялостно метагеномно секвениране бяха генерирани извън Qiime2, те бяха допълнително въведени в платформата преди да бъдат анализирани с приставката. Всеки отделен анализ беше съпроводен с различни параметри на филтриране в зависимост от обема на данните. Поради това конкретните параметри на филтриране бяха обозначени в описанията на фигурите.

3.5.6. Асемблиране и ко-асемблиране

Асемблиране и ко-асемблиране беше извършено посредством приложението megahit v1.2.9 с настройки по подразбиране, като бяха използвани *kmers* с дължина 21, 29, 39, 59, 79, 119, 141 и бяха филтрирани контиги по-къси от 1000bp. В някои случаи беше използвано приложението metaSPAdes v3.15.5, когато пробите бяха с по-малък обем, поради високите компютърни изисквания на това приложение, които не позволяваха да бъде завършен процеса. Цялостен анализ на проби МТС, преминали през цялостно метагеномно секвениране бяха анализирани за КНИ чрез приложението miqScoreShotgunPublic.

3.5.7. Биниране

Биниране беше извършено с три различни приложения: MaxBin2 v2.2.7, metaBAT2 и CONCOCT v1.1.0 поотделно като бяха използвани параметрите по подразбиране за всеки инструмент. Консолидирането на получените метагеном асемблирани геноми (МАГ) беше извършено с модула „bin re-finement“ на MetaWrap v1.3. След това получените асембли бяха оценени с приложенията QUAST v5.2 и CheckM v1.2.2 за качество, идентифицирани чрез GTDB-TK v2.3.2 и NCBI, анотирани и качени в Bioproject PRJNA1071831. Получените МАГ бяха подложени на таксономичен и функционален анализ.

3.5.8. Детекция на ГАР, ПГЕ и плазмиди

Суровите прочити бяха използвани за идентифициране на ГАР със софтуера ResFinder v2023-08-22 и базата данни v2023-04-12 или по-нова, като бяха приложени граници за 60% покритие и 60% идентичност, както и с приложението RGI v6.0.3 с базата данни CARD v3.2.8. ГАР от асемблита и МАГ бяха идентифицирани чрез Resfinder и AMRFinderPlus v3.11.14. ПГЕ бяха търсени в асемблитата посредством приложението MobileElementFinder v1.1.2.

Предполагаемите плазмиди бяха идентифицирани с помощта на три инструмента за прогнозиране на плазмиди: DeepPlasmid с минимална стойност на сигурност ≥ 0.75 , PlaSquid v1.0.0 и Plasme v1.1. Контигите, идентифицирани като плазмиди от поне един инструмент, бяха включени в анализа. Беше отчетена едновременната поява на ГАР или ПГЕ, или и на двете върху контиги с плазмиден произход.

3.5.9. Оценка на рисков фактор за човешко здраве

MetaCompare2.0 беше използван едновременно и за оценка на риска от мобилизируеми ГАР върху патогени [15].

3.5.10. Анализ на мрежите

Анализ на мрежите беше извършен с пакета NetworkX Python v3.2.1, като бяха включени само асоциации с положителна оценка на корелацията на Pearson ≥ 0.8 , съчетана с p -стойност ≤ 0.005 . След това мрежовата графика беше изготвена в Cytoscape 3.9.1. Статистическите изчисления бяха извършени чрез Scipy v1.11.1. Като входящи данни бяха зададени брой сурови прочити както на таксоните, така и на ГАР, като бяха запазени само такива с най-голяма представителност и такива, които се срещат във всичко проби, с изключение на гена *bla*_{IMP-13}-подобен.

4. РЕЗУЛТАТИ И ОБСЪЖДАНЕ

АМР и метагеномни проучвания са основни отправни точки в заглавието на тази дисертация. Въпреки че откриване и анализи на АМР е възможно само чрез цялостно метагеномно проучване, настоящият труд включва и ампликоново проучване, което обхваща стъпки на оптимизация и разработване на протоколи. Основната причина за това е, че до момента на дисертацията в Националната Референтна Лаборатория „Контрол и мониториране на антибиотичната резистентност“ не са извършвани метагеномни изследвания. С оглед на сравнително високата стойност на консумативите и реагентите за цялостни метагеномни изследвания и натрупване на опит, решихме да започнем с 16S ампликоново проучване като преданалитична подготовка като се фокусираме върху трите главни елемента: изолиране на ДНК, концентрация на биомаса, и биоинформатичен анализ.

4.1. Оптимизиране на *in-house* протокол за изолиране на ДНК

Един от основните подводни камъни в едно метагеномно проучване е изолирането на тотална геномна ДНК. В идеалния случай всички микроорганизми трябва да бъдат разрушени в равномерна степен спрямо тяхната реална концентрация в пробата и да се получи ДНК от тях в пропорционални съотношения. Въпреки това, стъпката на изолиране на ДНК внася известни отклонения като разрушава по-успешно само някои микроорганизми. Поради композиционната структура на метагеномните данни дори малки изменения в изолиране на ДНК от един микроорганизъм в съобществото води неминуемо до изменения в останалите представители в крайния резултат. Това изисква балансирано лизиране на всички целеви микроорганизми. Към такъв резултат се стремят всички разработчици на комерсиални и *in-house* протоколи за изолиране на ДНК [6].

Поради тази причина и други като цена, достъпност и надеждност в настоящия труд беше заложена задачата за изграждане на протокол за изолиране на ДНК за метагеномно проучване като се разчита на вече утвърдени добри практики. Бяха използвани моделни бактерии с трудни за лизиране клетъчни стени: *S. aureus* ATCC 25923; *C. Diphtheriae*; *B. Licheniformis*; *E. faecalis* и *E. faecium*, върху които да се изпитат разнообразни техники за лизиране. Разнообразие от реактиви, лизиращи буфери, подходи и техники бяха изпитани в различни комбинации като: механичен лизис с перли (стъклени и метални) с различни размери в присъствието на различни лизиращи буфери; биологичен лизис с лизозим при различни условия; физичен лизис с ултразвук; химичен лизис с детергенти в различни концентрации и комбинации в лизиращи буфери. Оценяването се извършваше чрез количество добив на ДНК, измерена флуориметрично, спектрофотометрично (съотношение на 260/280 и 260/230), както и електрофоретично. Финализиращият протокол беше подложен на сравнителен анализ. За целта проба МТС беше подложена на изолиране на ДНК с *in-house* протокола, както и четири други комерсиално използвани кита за изолиране на ДНК.

4.2. Дизайн за изпитване на ФОМ и ВФ

Независимо от методите, дизайна и целите, пробите от различни водни обекти се събират, транспортират до лабораториите и се обработват незабавно, за да се избегнат необективни резултати. Началната стъпка на повечето бактериални метагеномни изследвания върху водни проби включва извличане на биомаса (наричано още обогатяване на биомасата) от водното съдържание. Под биомасата се визира целевите микроорганизми, които са в планктонни форми или адхерирани върху малки частици като почва, пръст, растителни остатъци, и ситни камъчета или пясък. За да се улови цялото микробно съдържание, в идеалния случай тези частици трябва да се задържат в крайната проба. За тази цел са разработени и използвани в различна степен няколко техники, а именно ВФ, ФОМ, ултрацентрифугиране и утаяване с полиетиленгликол. Последните два се прилагат почти единствено за вирусна и фагова метагеномика.

ВФ се използва главно за характеризиране на бактериални и гъбични общности и досега се счита за златен стандарт. Поради множество причини изследванията върху методите за филтриране все още се развиват и се търсят алтернативни подходи. От друга страна, ФОМ може да се счита за алтернативна техника за обогатяване с бактерии от водните проби. Първоначално тя е разработена за концентриране на вируси от крайбрежни води [16], а по-късно е широко прилагана за различни вируси включително и SARS-CoV-2 [17]. Тя е лесна, достъпна и бърза за изпълнение и изисква просто лабораторно оборудване. По-рано ФОМ е използвана за едновременно концентриране и количествено определяне на вируси, бактерии и протозои, пренасяни с водата [18], главно за целите на контрола на водите и проучвания за оценка на микробния риск. В литературата, обаче, липсва информация за приложимостта му при изследвания на бактериалния микробиом. Поради това в настоящия труд беше поставена задачата да се оцени приложимостта на ФОМ за бактериална метагеномика при водни проби, като се съпостави спрямо ВФ и по двете техники за метагеномика.

4.3. Резултати от сравнителни анализи върху *in-house* протокола, 16S-БД и ФОМ

Първо беше анализирано отклонението, въведено от различните протоколи за извличане на ДНК и 16S-БД върху данни от пробите МТС ($n=8$). Бяха оценявани фактори като добив на ДНК A260/280nm, A260/230nm, TOT и ЧОТ на нива род и вид, КНИ, процент на нуклеотидни прочити, които се съпоставят с референтните секвенции, такива, които не са преминали филтъра за качество, такива, които не са се събрали (прав и обратен) или са били химерни (Виж *Таблица №11*). *In-house* протоколът имаше най-висок добив на ДНК, докато най-висока чистота беше постигната с китове EurX и EZNA.

Таблица №11. Първоначални данни за пробите МТС. Резултатите за КНИ стойности и оценяване на брой прочити, преминали анализа е извършен с оригиналното приложение miqScore16SPublic.

Кит/ протокол	ДНК ng/µl	260/ 280	260/ 230	КНИ	Прочити отг. на реф, %	прочити с ниско качество,	Несдвоени прочити, %	Химерни прочити, %
Zymo-55°C	27	1.92	0.28	65	36.6	25.2	30.3	8.0
Zymo-58.5°C	27	1.92	0.28	76	40.0	24.1	12.0	23.9
Zymo-62°C	27	1.92	0.28	65	52.3	22.0	12.1	13.6
Zymo-62°C-2	27	1.92	0.28	71	46.7	23.0	26.4	3.9
EurX	28.5	1.96	1.16	79	53.9	22.0	13.3	10.8
EZNA	32	2.13	2.28	74	61.6	21.8	15.9	0.7
In-house	63	1.75	0.66	78	57.7	25.4	13.3	3.7
FastSpin	47	1.88	0.14	88	48.4	21.4	13.1	17.0

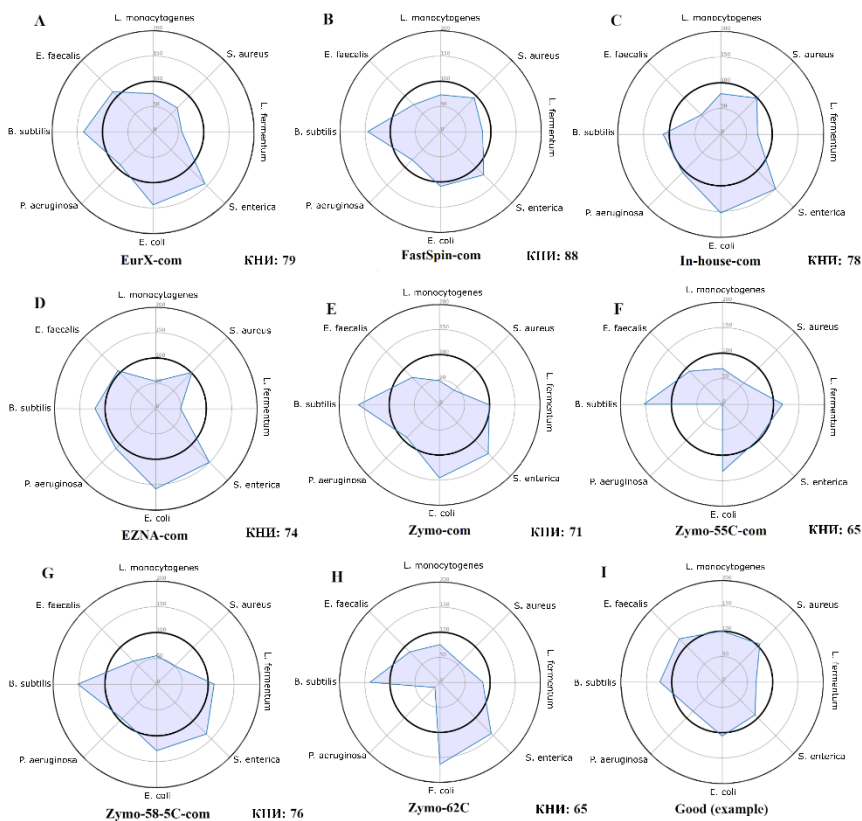
4.3.1. КНИ стойности и отклонения от протоколите за изолиране на ДНК

Приложението miqScore16SPublic, предоставено от Zymo, беше използвано за установяване на отклонението, въведено от протоколите за екстракция на ДНК. Радарните графики заедно с КНИ стойностите са представени на *Фигура №3*. Резултатите показват, че пробите, изолирани с комплекта FastSpin Soil, показват най-малък резултат на отклонение (88 КНИ), следвани от EurX и *in-house* протокола, докато най-лошите резултати на КНИ са в комплекта Zymo и неговите варианти с различни температури на хибридизация.

Тези резултати бяха генерирани първи в хронологичен аспект и въз основа на тях и резултатите от *Таблица №11* беше взето решение, че китът EurX предоставя относително добри резултати по всички категории и е пригоден за водни проби. Поради това той беше избран за изолиране на ДНК от пробите за цялостен метагеномен анализ.

4.3.2. Отклонения от 16S-БД и/или наборите за изолиране на ДНК върху проби МТС

Въпреки че научните изследвания постепенно изместват фокуса си към цялостното метагеномно секвениране, множество усилия се полагат върху стандартизирането на 16S-БД и протоколите за правилна таксономична оценка. Съществуват различни подходи за клъстериране при обработване на секвенционни данни от ампликоново метагеномно проучване като най-често ползваните са *open*- или *closed-reference* ОТЕ клъстериране или идентифициране на АСВ. Въпреки че най-надеждната идентификация обикновено се постига с АСВ, понеже тя е биологически смислена единица, голяма част от проучванията все още разчитат на *open*- или *closed-reference* ОТЕ клъстериране, макар и на цената на отклонения, внесени от самите 16S-



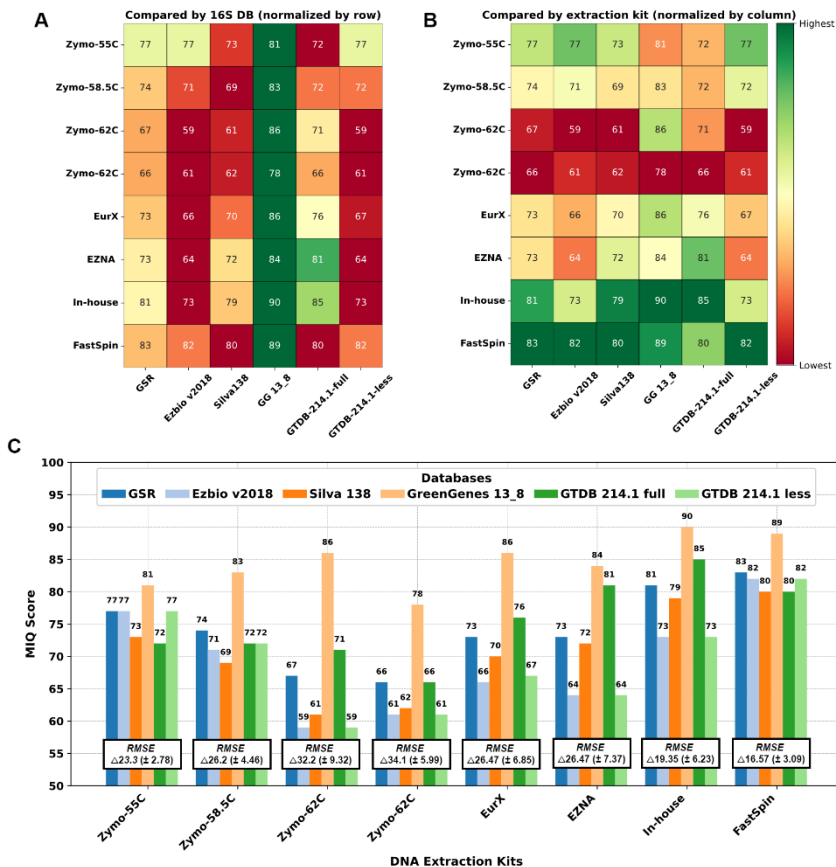
Фигура №3. Радарни диаграми на всички проби МТС (А-Н), оценени с помощта на оригиналното приложение miqScore16SPublic. Тези диаграми показват наблюдаваното съотношение на всеки организъм спрямо очакваната му стойност. В идеалния случай всички точки трябва да бъдат разположени около марката 100% (вътрешния удебелен кръг), което показва, че наблюдаваните пропорции са на нивото на очакваните стойности или са близки до тях, както е показано на диаграма I (МТС с добра оценка и минимални отклонения). Общото отклонение във всяка извадка се изчислява в общ показател КНИ (по-високата стойност е по-добре), като резултатите >90 се считат за отлични, а $80 \div 89$ - добри.

БД. *De novo* клъстерирането често е първи избор за наблюдение, тъй като съставът не се влияе от 16S-БД, но той не е оптимален за междублабораторни сравнения. Поради това, *closed-reference* ОТЕ клъстериране е може би най-широко прилаганият подход, но пък е пряко повлиян от 16S-БД. Нови БД биват публикувани често, а други актуализирани, което ги прави обект на множество проучвания и сравнителни анализи за тяхната ефективност и приложимост [19].

За оценяване степента на отклонение от 16S-БД в таксономичния състав беше приложено 99% *closed-reference* ОТЕ клъстериране на пробите МТС. За съжаление инструментът *miqScore16SPublic* беше несъвместим с таксономични таблици, генерирани от външни източници. Като заобиколен вариант беше приложен неговата вътрешна формула за КНИ в Python скрипт, което позволява изчисляването да се прилага към таксономични таблици независимо от техния произход. По този начин беше изчислен КНИ на всяка ОТЕ клъстерирана проба. Графиките на таксономичния състав на всички проби МТС, както и техните КНИ стойности са визуализирани на *Фигура №4* и *Фигура №5*.

Фигура №4. Бар диаграми на таксоните на осемте проби МТС, анализирани със *closed-reference* клъстериране с всяка 16S-БД поотделно, както и на наборите от данни с *de novo* клъстериране, идентифицирани с GTDB-full. В дясната страна са представени най-нисък ранг таксони, специфични за всяк 16S-БД. По-високият резултат на КНИ представлява таксономичен състав, по-близък до теоретично очаквания (най-долу).

16S-БД GTDB-full и Silva генерират повече ОТЕ в сравнение с останалите 16S-БД. Това може да се счита както за добър, така и за лош резултат в зависимост от целта на изследването. В този случай, допълнителният таксон *Citrobacter* в GTDB-full, който не би трябвало да присъства в пробата МТС, изкривява резултатите, което води до по-ниски КНИ стойности (Виж *Фигура №4*).



Фигура №5. Ефективност на наборите за изолиране на ДНК и 16S базите данни по КНИ стойности. **A.** Сравнение на 16S бази данни. **B.** Сравнение на наборите за изолиране на ДНК/температури на хибридазация. Зеленото е по-добро, а червеното - по-лошо. **B.** Бар диаграми с оценките на КНИ и средните стойности на root mean square error (RMSE).

Успоредно бяха съпоставени КНИ стойностите спрямо 16S-БД и протоколи за ДНК изолиране (Виж *Фигура №5*). Първо бяха сравнени 16S-БД (Виж *Фигура №5A*) и най-близките до очакваните таксономични състави с най-малко отклонение бяха

получени от GG_13.8, следвани от GSR и GTDB-full. Интересно е, че GG_13.8 БД не успява да разграничи *Escherichia coli* и *Salmonella enterica*, като ги обединява в една група ОТЕ на ниво семейство *Enterobacteriaceae* (Виж *Фигура №4*). За съжаление, няма надежден подход за разделяне на тази 99% ОТЕ група, за да може КНИ калкулатора да определи точно отклонението за всеки микроорганизъм по отделно. Ето защо конкретно за GG_13.8 третирахме *E. coli* и *S. enterica* като един организъм и коригирахме референтния очакван състав в калкулацията на КНИ. В резултат на това всички КНИ стойностите от GG_13.8 бяха повишени в сравнение с останалите 16S-БД. Поради това тези стойности не са достоверен индикатор за ефективността на GG_13.8 БД и трябва да се тълкуват с повишено внимание.

След това бяха сравнени ефективността на различните китове за екстракция на ДНК (Виж *Фигура №5B*), независимо от използваните 16S-БД. С изключение на ZymoBIOMICS с температура на хибридизация 55°C, всички варианти на Zymo се представиха по-зле в сравнение с другите протоколи. Китът FastSpin постигна най-високи резултати по КНИ (средно 82.6) с всички 16S-БД. Вторият най-добър резултат беше постигнат с нашия *in-house* протокол.

Интересно беше, че всички 16S-БД освен GG_13.8 постигнаха близки резултати при всички протоколи за изолиране на ДНК, като КНИ варираше в диапазона 59 ÷ 71 при най-лошо представящата се (Zymo-62C) и 80 ÷ 83 при най-добрата проба (FastSpin). Тези резултати предполагаха, че изборът на протокол за ДНК изолиране е оказал по-голямо влияние върху крайния таксономичен състав на МТС. Наблюдаваното отклонение в пробите МТС не може да бъде компенсирано чрез използване на по-добре представяща се 16S-БД (както се вижда при двете повторения на Zymo-62C на *Фигура №5C*). И обратно - проба, обработена с добре работещ протокол за ДНК изолиране, предоставя таксономичен състав близък до очаквания независимо от използваната 16S-БД.

Извличането на ДНК е критична стъпка в работния процес на метагеномиката и е известно, че се влияе от множество параметри, което прави цялостно й оценяване предизвикателство. Освен това, изборът на метод за изолиране на ДНК оказва силно влияние върху откриването и състава на бактериалните съобщества [20]. Непрекъснато се разработват и прилагат некомерсиални методи и протоколи и/или търговски продукти за тази цел, което води до много разнообразни резултати. Разбира се, няма перфектен протокол за изолиране на ДНК, който да не внася отклонения, но тези отклонения могат да бъдат драстични както сочат резултатите от настоящия труд. А съставът, разнообразието и разпределението на таксономичните данни от едно проучване са крайната цел и обект на сравнение между проучванията. Въпреки че разработихме добре работещ протокол за изолиране на ДНК, са необходими допълнителни подобрения, за да се постигне съответствие с ефективността на комплекта EurX (с ACB) или FastSpin (с ACB и ОТЕ). Изненадващо беше, че китът EZNA Universal Pathogen не е предназначен за метагеномика и със сигурност не е

оптимизиран да извлича ДНК еднакво от Грам-положителни и Грам-отрицателни бактерии, тъй като клетъчната стена на Грам-положителните бактерии съдържа дебел слой пептидогликан. Въпреки това той предостави близки по КНИ стойности резултати като други набори като Zymo-BIOMICS и EurX, дори без включва стъпка на механичен лизис с перли, която понастоящем е широко приета практика и се препоръчва за балансиран лизис.

4.3.3. Ефикасност на идентифициране на 16S базите данни при *de-novo* клъстериране

Докато таксономичният състав е първият критерий за оценка на отклонението в пробите МТС, TOT и ЧОТ са два важни фактора, които оценяват как 16S-БД извършват таксономичната идентификация. Въпреки че резултати са представени като характеристики и оценка за всяка 16S-БД, трябва да се отбележи, че те са силно повлияни от избора на амплифицирания регион от таргетния ген 16S и избора на формиране на таксономични единици (ОТЕ спрямо ACB) и не трябва да се разглеждат изцяло като недостатъци на 16S-БД.

Резултатите от TOT и ЧОТ са изчислени с данни от 99% *de novo* клъстерирани ОТЕ групи, тъй като таксономичният състав на всички проби МТС е един и същ, независимо от 16S-БД, което осигурява еднаква база за сравнение. Резултатите показват, че *de novo* клъстерирани ОТЕ на V3-V4 региона с тези праймерни двойки са склонни да подценяват *Enterococcus faecalis*, като го разделят на два отделни таксона *Enterococcus* spp. и *Enterococcus faecalis* със сходни стойности на относителната честота във всички набори за екстракция, като по този начин изкривяват общия микробен състав. За сравнително малък МТС със само 8 бактериални вида TOT и ЧОТ не се различаваха драстично между отделните протоколи за екстракция на ниво род и вид бяха идентични. На ниво вид TOT варираше от 1/8 бактерии за Silva, 2/8 за GSR и до 4/8 за Ezbio, докато за родово ниво тя беше между 6/8 и 8/8 (Виж Таблица №12 и Фигура №4). GG_13.8, GTDB-full и GTDB-less са предназначени за идентификация на ниво род, поради което идентификация до ниво вид не е възможно и не беше обсъждано. Беше регистриран само един случай на неправилна класификация, а именно *Listeria monocytogenes*, идентифицирана като *Listeria ivanovii* от Ezbio. Недостатъчно класифициране с две таксономични нива се наблюдава за *Salmonella enterica* при GTDB-full, GTDB-less и Silva и за *Escherichia coli* по GG_13.8 и GSR. Освен това *P. aeruginosa* е класифицирана с две степени по-високо по GG_13.8. Таксономичният състав на *de novo* клъстерираните ОТЕ проби се вижда на „*de novo* + GTDB-full“ на Фигура №4.

За сравнението на 16S-БД беше извършена възможно най-индивидуализираната таксономична идентификация за настоящата МТС проба, двойката праймери и за 16S-БД. Това беше извършено като първо бяха изведени само регионите на амплификация от целите последователности на референтните гени 16S и използвани за конструиране на класификатор. По този начин TOT и ЧОТ на всяка

16S-БД е специално адаптирана към използваната двойка праймери, което позволява стандартизирано сравнение. Най-големият фактор за внесени отклонения на ниво биоинформатична обработка се явяват БД, тяхната пълнота и колко валидирани таргетни гени включват. Поради това се излага становище, че вместо големи 16S-БД, включващи цялото познато микробно разнообразие, ще се премине към специфични за всяка екосистемата БД.

Таблица №12. Точност на определяне и честота на откриване на таксони от 16S-БД при МТС. Данните са представени в процент от всички микроорганизми ($n=8$).

Кит/ протокол	TOT/ЧОТ на ниво вид, %					TOT/ЧОТ на ниво род, %				
	GSR	Ezbio v2018	Silva 138	Green Genes 13_8	GTDB- 214.1- full & less	GSR	Ezbio v2018	Silva 138	Green Genes 13_8	GTDB- 214.1- full & less
Zymo-55°C	25	50	12.5	0	0	87.5	100	87.5	75	87.5
Zymo-58.5°C	25	50	12.5	0	0	87.5	100	87.5	75	87.5
Zymo-62°C	25	50	12.5	0	0	87.5	100	87.5	75	87.5
Zymo-62°C-r2	25	50	12.5	0	0	87.5	100	87.5	75	87.5
EurX	25	50	12.5	0	0	87.5	100	87.5	75	87.5
EZNA	25	50	12.5	0	0	87.5	100	87.5	75	87.5
In-house	25	50	12.5	0	0	87.5	100	87.5	75	87.5
Fast Spin	25	50	12.5	0	0	87.5	100	87.5	75	87.5

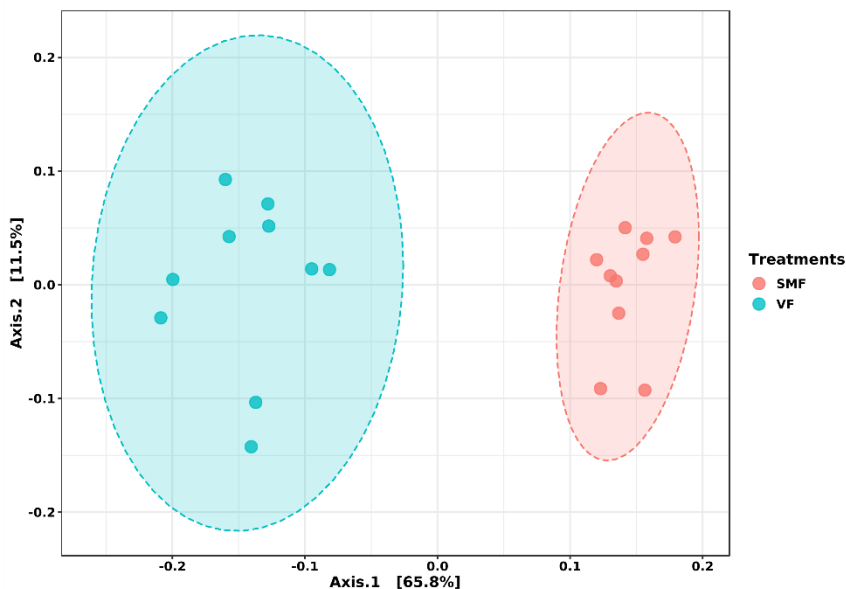
Стойностите на TOT и ЧОТ не бяха толкова информативни, както се очакваше първоначално, главно поради малкия брой включени бактерии ($n=8$). Представените резултати от сравнението на 16S-БД трябва да се интерпретират заедно с амплифицирания регион V3-V4, тъй като идентификацията също е силно повлияна от този фактор. TOT на получените ОТЕ групи, клъстерирани при граница, равна на 99% или по-ниска, обикновено страдат от отклонение в идентификацията. Както се очакваше не беше постигнат 8/8 TOT с ОТЕ.

И не на последно място, оптимизираният протокол за генериране на 16S секвенционни библиотеки, описан в настоящия труд и приложен в ампликоновото проучване, показва добри КНИ резултати (≥ 80) с повечето китове за изолиране на ДНК и *in-house* метода. Според интерпретацията на КНИ от оригиналното приложение miqScore16SPublic тези резултати гарантират, че ако една проба МТС е приложена по предназначение като контрола на ДНК изолирането, то КНИ стойности по-големи от 80 биха валидирали резултатите от реалните проби. Въпреки това има място за допълнителни подобрения в протокола. За да се оцени напълно приложимостта на протокола за 16S библиотеки са необходими допълнителни сравнителни анализи с други протоколи за генериране на 16S библиотеки.

4.3.4. Оценка на флокулацията с обезмаслено мляко

ФОМ и ВФ бяха сравнени само върху реални проби тъй като МТС от Zymo не е предназначен за предварителна обработка преди извличане на ДНК, тъй като

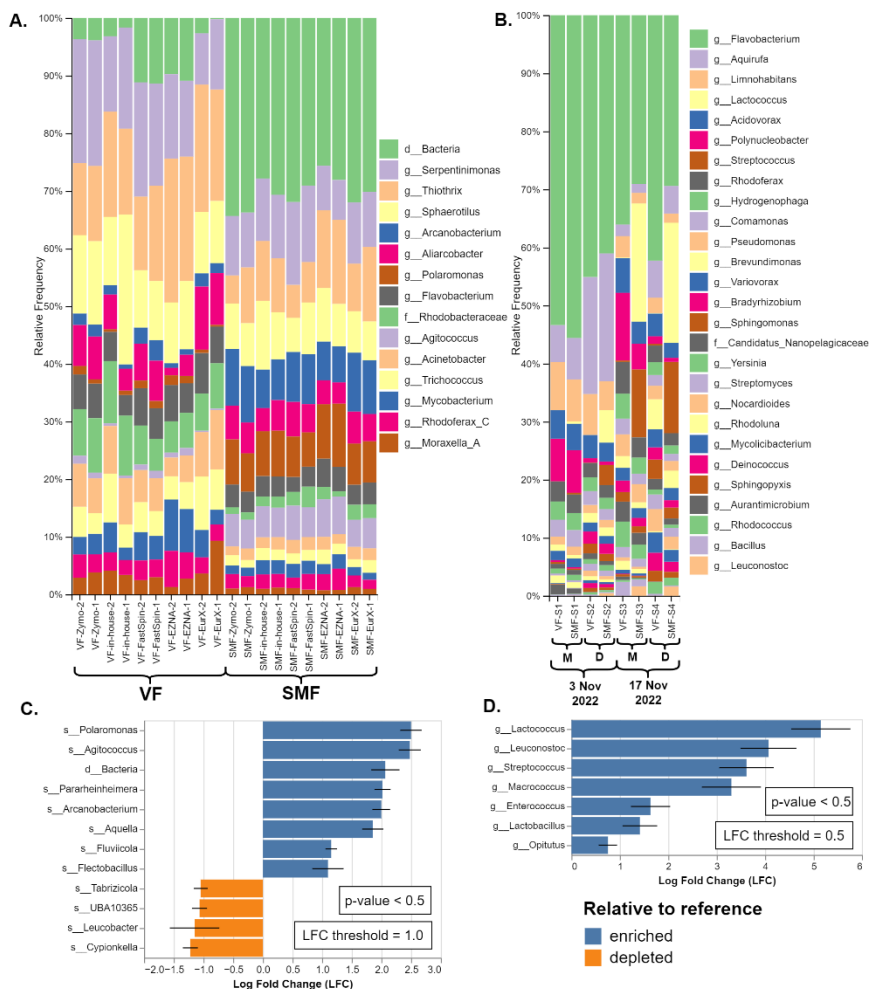
клетките се съхраняват в DNA/RNA Shield™ и са частично лизирани. За по-пълноценно съпоставяне на методиката с ВФ, бяха използвани успоредно ампликоново и цялостно метагеномно проучване. По този начин, чрез комбиниране на двата вида методики, успешно бяха разгледани промените в бактериалния видо-состав между проби, третирани с двете техники. Още при първоначалните анализи на бета разнообразието стана ясно, че двете групи проби от 16S ампликоновото секвениране бяха ясно разделени на АГК графиката (PERMANOVA: $F = 26.6$, $R^2 = 0.6$, $p < 0.001$), представена на *Фигура №6*.



Фигура №6. Анализ на главните компоненти показващ разделянето на обработените проби с ФОМ и ВФ чрез 16S секвениране от река Перловска. Графиката е изготвена с ResistoXplorer чрез филтриране на таксони с нисък брой (<5) с разпространение в поне 4 проби и прилагане на скалиране чрез обща сума нормализиране. [PERMANOVA] F-стойност: 26.655; R^2 : 0.59691; $p < 0.001$.

За да бъдат оценени кои микроорганизми са движещите фактори, водещи до тези значими разлики между пробите, третирани по двете техники за концентриране, беше използван АДИ. Този анализ използва подходи за скалиране, предимно базирани на логаритмуване на данните, за да определи дали даден таксон е различен (преувеличен или намален) между две проби или две групи от проби. По литературни данни приложението ANCOM-BC2 беше представен като най-добър кандидат [21], който може да се справя с проби както с ниско, така и с високо биологично

разнообразие. Получените профили на относителния таксономичен състав на двата вида проби са представени на *Фигура №7А* и *Фигура №7В*.



Фигура №7. Сравнение на таксономичния състав на ФОМ и ВФ. **А.** Двойки проби от река Перловска, подредени по набор за извличане на ДНК (ампликонова метагеномика). **В.** Набори от данни от пробите за цялостно метагеномино секвениране от река Искър. Двойките проби (третираны с ВФ и ФОМ) са подредени по дата на вземане на пробата и местоположение (М - вилна зона Мечката, Г - с. Драгушиново). Анализ на диференциалното изобилие на **С.** ампликонните и **Д.** shotgun пробите. Всички ВФ проби бяха обединени като обща референтна група, и съпоставена с групата на ФОМ-третираны проби. Представени са само таксони със стойност $p < 0.5$ и с ЛПП ≥ 1.0 (**С.**) и 0.5 (**Д.**).

В проби, третирани с ФОМ, се наблюдава свръхпредставяне на родовете *Polaromonas*, *Agitococcus* и една ОТЕ група, идентифицирани с висок таксономичен

ранг. Род *Polaromonas* беше идентифициран чрез GG_13.8 и GSR на видово ниво като *P. naphthalenivorans*, докато ОТЕ групата беше или домейн *Bacteria*, разред *Bacteroidales*, или семейство *Saprospiraceae* от EzbioCloud (както безплатната БД v2018, така и най-актуалната онлайн БД v2023.08.23). И трите таксона заедно с още няколко се оказаха значително обогатени ($p < 0.5$) с логаритмична промяна в пъти (ЛПП, от англ. „log fold change – LFC“) стойност от 2.0 или повече във всички китове за извличане на ДНК, както е показано от DAA на *Фигура №7C*.

Подобна тенденция се наблюдаваше и при цялостното метагеномно секвениране, но с различен профил на обогатените таксони (Виж *Фигура №7D*). Не всички обогатени таксони са видими на бар графиките, с изключение на родовете *Streptococcus* (кафяво) и *Lactococcus* (бледожълто) в пробите SMF-S2, SMF-S3 и SMF-S4 на *Фигура №7B*. Според DAA членовете на млечнокиселите бактерии *Lactococcus*, *Leuconostoc*, *Streptococcus*, *Enterococcus* и *Lactobacillus* са значително обогатени ($p < 0.5$) с ЛПП между $1.75 \div 5.0$ в пробите, третираните с ФОМ. Род *Macroccoccus* също беше значително увеличен, а идентифицираните нуклеотидни прочити на ниво вид от приложението Bracken бяха предимно *Macroccoccus caseolyticus*, за който отново е доказано, че ефективно хидролизират казеини и е естествен компонент на вторичната микрофлора в сирена и колбаси. Интересно беше, че родът *Streptococcus* беше открит също в ампликоновите метагеномни проби, но не беше свръхпредставен.

Въпреки че профилите на тези обогатени таксони варираха при различните проби и методики, ефектът на обогатяване беше ясно отразен и оказва значително влияние върху крайния таксономичен състав. Ефектът беше консистентен въпреки различията в протоколите за извличане на ДНК (16S ампликоново метагеномно), и различните дати на вземане на проби (цялостно метагеномно). Следователно този ефект се проявяваше независимо от методологията или вида на пробата.

Най-логичната и подкрепена с доказателства от литературата беше хипотезата, че обезмасленото мляко е било метаболизирано по време на двучасовия инкубационен протокол, което е довело до размножаването на специфични таксони. Обезмасленото мляко се състои предимно от белтъчини (казеин и суроватка), лактоза и други хранителни вещества и минерали, които действат като растежни фактори за бактериите и обикновено се добавят културелни среди за микробиология. В тази връзка е било доказано, че род *Polaromonas* е обогатен в резервоарите за отстраняване на млечни продукти, докато други проучвания го определят като третия най-разпространен род в биофилм от смесени видове млечни продукти в биофилтри [22]. Родът *Agitococcus* също беше значително обогатен от ФОМ. Въпреки че не беше постигната идентификация на видово ниво, видът *Agitococcus lubricus*, описан през 1981 г., е бил положителен за протеолиза на обезмаслено мляко. Вероятно и други представители на род *Agitococcus* са способни на протеолиза.

По подобен начин и при другата методика всички обогатени млечнокисели бактерии обикновено се откриват в разлагащи се растения и млечни продукти,

произвеждайки млечна киселина като основен метаболитен краен продукт на въглехидратната ферментация чрез оползотворяване на лактозата от обезмасленото мляко. Подкиселяването на пробите ($\text{pH}=3.5$) по време на ФОМ, което опосредства процеса флокуляция, може да е било благоприятно за тяхното възпроизвеждане. И накрая, според таксономичните таблици на Bracken, родът *Macrococcus* най-вероятно е бил представен основно от *Macrococcus caseolyticus*,

4.4. Анализи, базирани на цялостни метагеномни проучвания

4.4.1. Сурови данни и МТС

Пречистените сурови прочити варираха от 41.6 до 63.8 млн и 2.20 млн за МТС за пробите от р. Искър. Средната дължина на прочитите за проба варираше между 142 и 150 нб. За пробите от гр. Стара Загора прочитите варираха от 61.8 до 123.4 млн.

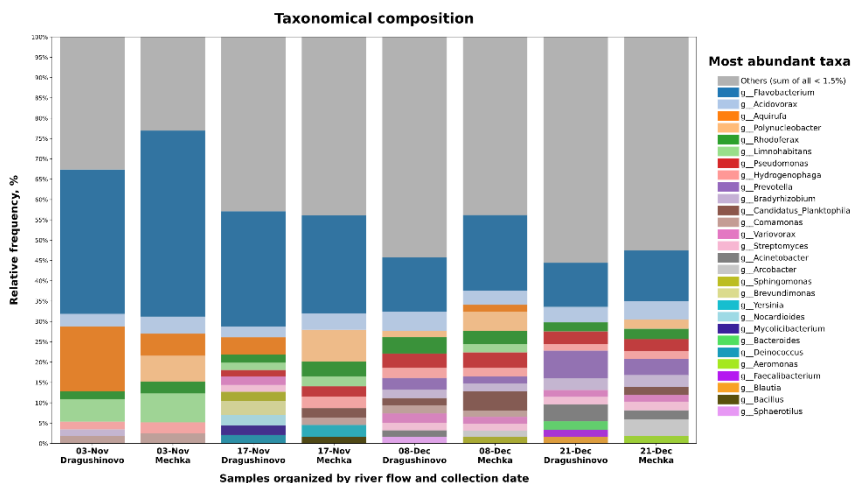
Заложените като контроли на ДНК изолирането проби МТС бяха обработени отделно. КНИ на МТС при пробите от р. Искър възлезе на 67 точки, което означава, че се наблюдават нива на отклонение, внесени в настоящия анализ от изолирането на ДНК и секвенирането. МТС за сондажните води и двете ПСОВ от гр. Стара Загора възлязоха на 81.

4.4.2. Влиянието на ПСОВ Самоков върху р. Искър

Река Искър тече изцяло на територията на България. Средният годишен дебит на река Искър е $54.5 \text{ м}^3/\text{с}$. Тя е основният източник на вода за близо 80% от домакинствата (около 1.3 млн души) и промишлените нужди на столицата София и Софийска област. На първо място тя захранва община Самоков, където се намира ПСОВ гр. Самоков, която обхваща и обслужва площ от 125,000 еквивалент жители, има максимален капацитет от 16,060,000 $\text{м}^3/\text{г.}$, но пречиства само на базата на механично и биологично пречистване (първично и вторично ниво), но няма третично пречистване. Тя е въведена в експлоатация през 2001 г., и е най-голямата ПСОВ без модернизация до момента, и се експлоатира от дружество „Софийска вода“. Според проучване от 2019 г. се наблюдава отрицателна асоциация между изобилието на ГАР в рециклираните води след ПСОВ и броя на приложените степени на биологично пречистване [23]. Пречистената вода се зауства в река Искър, като по този начин се превръща в първият голям източник на замърсяване, разположен по течението на реката. Само на 10 км по-надолу по течението река Искър се влива и захранва изкуствено изградения язовир "Искър" с обем 675 млн м^3 , който е популярно място за водни развлечения, а също така е основен източник на водоснабдяване и производство на електроенергия за гр. София и околните населени места. По този начин река Искър в този регион е от стратегическо икономическо значение. Понастоящем няма публикувани таксономични или данни за АМР за река Искър или региона. Тази причина беше основополагаща в избора на локациите за пробонабиране от този сладководен воден източник като ключов обект.

4.4.2.1. Таксономичен състав и разпространение

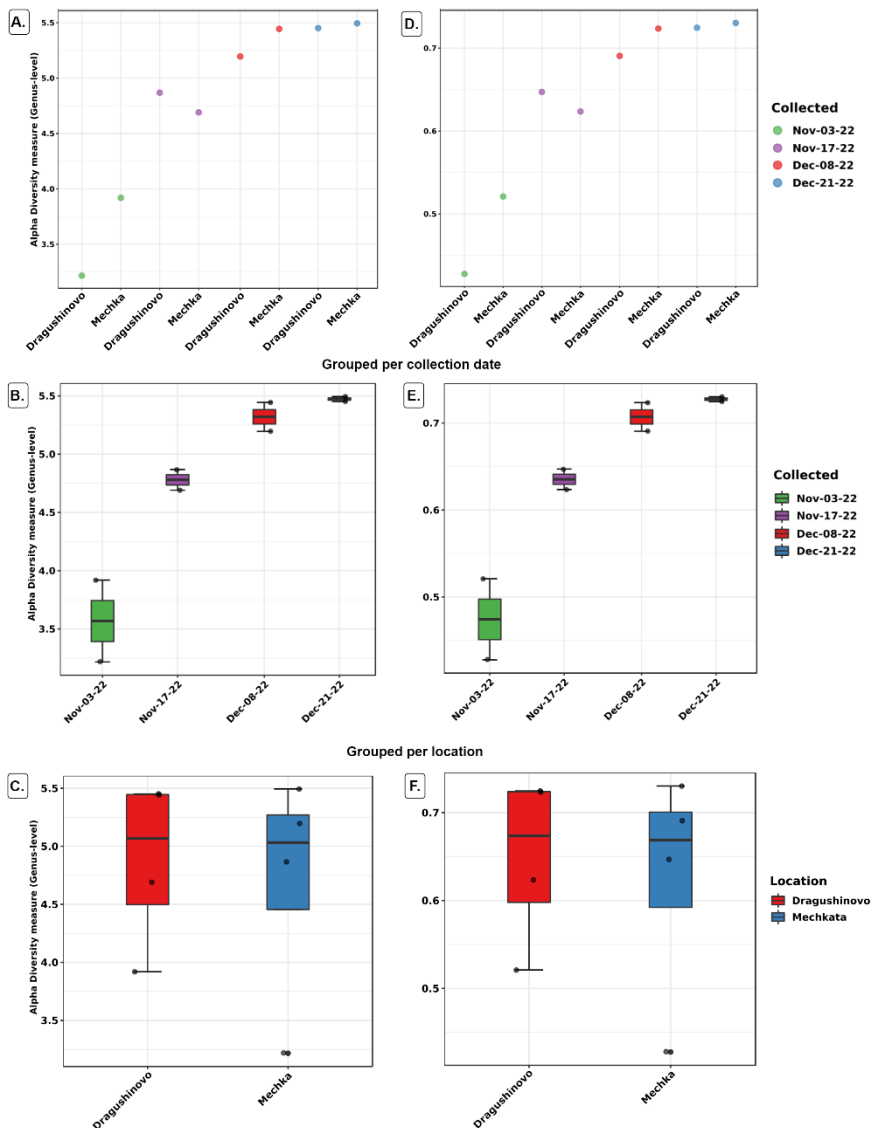
Таксономичния състав на ниво род на пробите е визуализиран на *Фигура №8*. Основните таксони, идентифицирани в пробите, бяха предимно *Pseudomonodota*, съставляващи 38% (при първите дати на пробовземане) до 60.7% (при по-късните дати), и *Bacteriodota*, показващи постепенно намаляване от 54% до 21%. В рамките на *Pseudomonodota*, *Betaproteobacteria* представляваха 28.6% на ниво клас, а *Burkholderiales* - над 95% на ниво разред. Сред тях родовете *Acidovorax*, *Polynucleobacter*, *Limnohabitans* и *Rhodoferrax* показаха най-високи относителни честоти и според литературните данни се срещат заедно в рамките на едно и също местообитание [24]. Най-разпространените таксони във всички проби включваха *Flavobacterium*, *Aquirufa*, *Acidovorax*, *Polynucleobacter* и *Limnohabitans*. Освен това род *Flavobacterium* преобладаваше в рамките на тип *Bacteriodota* съставлявайки 70 ÷ 73% от първоначалните дати, намалявайки до 60.6% на третата дата, и намалявайки още повече до 43.8% в края на периода на изследване.



Фигура №8. Барплот графика с таксономичен състав, изобразяващ относителното изобилие на най-разпространените родови таксони във всяка от осемте проби, подредени по дата и течение на реката. Нискоразпространените таксони (< 0.1%) и таксоните, открити в по-малко от 3 проби, бяха филтрирани. Броят на ниво род беше трансформиран в относително обилие, а таксоните под 1.5% бяха групирани в група "Други". Първите 30 рода са обозначени и оцветени в различни цветове и подредени от най-многобройните в низходящ ред въз основа на средното относително изобилие, изчислено за всички проби.

Наблюдавайки състава в динамика беше установено, че някои родове са склонни да намаляват с времето. Например, както *Flavobacterium* така и *Aquirufa* намаляват за период от седем седмици, докато други като *Acidovorax* показват постоянни нива във всички проби. Освен това *Polynucleobacter* е значително по-

високо представен на втората локация (вилна зона "Мечката"), разположена надолу по течението. Тези тенденции бяха последователни през четирите дати, което предполагаше, че промените между двете местоположения не се дължат на случайни събития и не са повлияни изцяло от времевия интервал. За сравняване на пробите поотделно или по групи бяха използвани показатели за алфа разнообразие на Shannon (Виж *Фигура №9A,B,C*) и Pielou (Виж *Фигура №9D,E,F*).



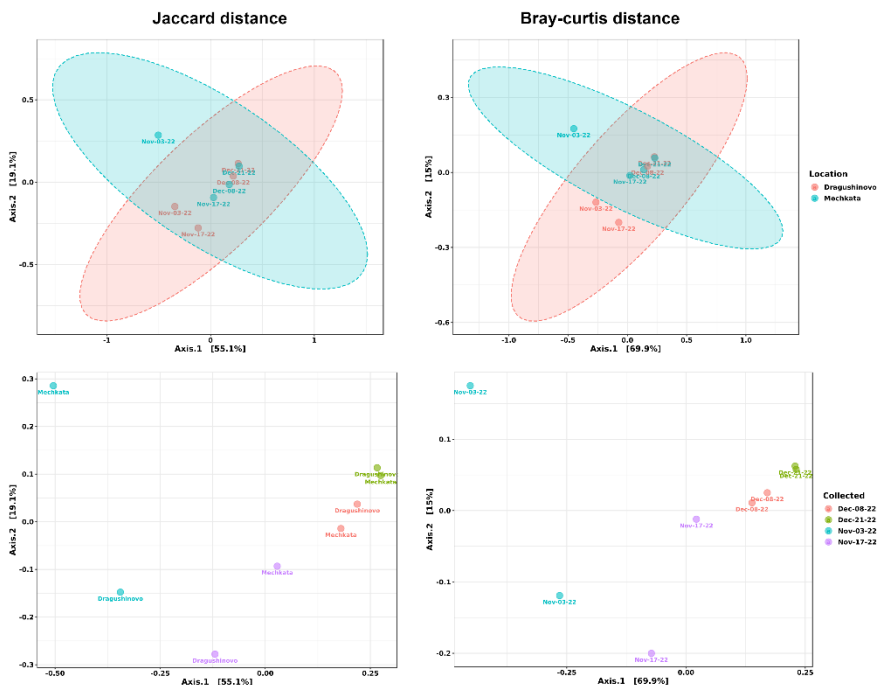
Фигура №9. Алфа разнообразие и тестване на значимостта на пробите по групи. Графиките показват по оста у мерките за алфа разнообразие (на ниво род) на Shannon (A, B, C) и Pielou (D, E, F). A. и D. всяка проба поотделно; B. и E. в групи по дата; C. и F. в групи по местоположение.

И двата индекса показаха нарастване на богатството и по-равномерно разпределение на видовете с течение на времето. Не беше установена разлика между двете групи според фактора местоположение (Mann-Whitney = 8, $p = 1$). И двата индекса, обаче, показаха незначителна тенденция на сближаване (Kruskal-Wallis = 6.6667, $p = 0.08$) между двойките проби от една и съща дата в хода на изследването. Най-голямото разминаване в индексите беше наблюдавано в началната дата, последвано от постепенно намаляване на разликите на следващите дати. Това до голяма степен беше повлияно от намаляването на рода *Flavobacterium*. Същият тип анализ беше проведен с таксономични таблици на ниво семейство, разред и клас и резултатите не се промениха.

4.4.2.2. Анализ на главните координати и анализ на диференциалното изобилие

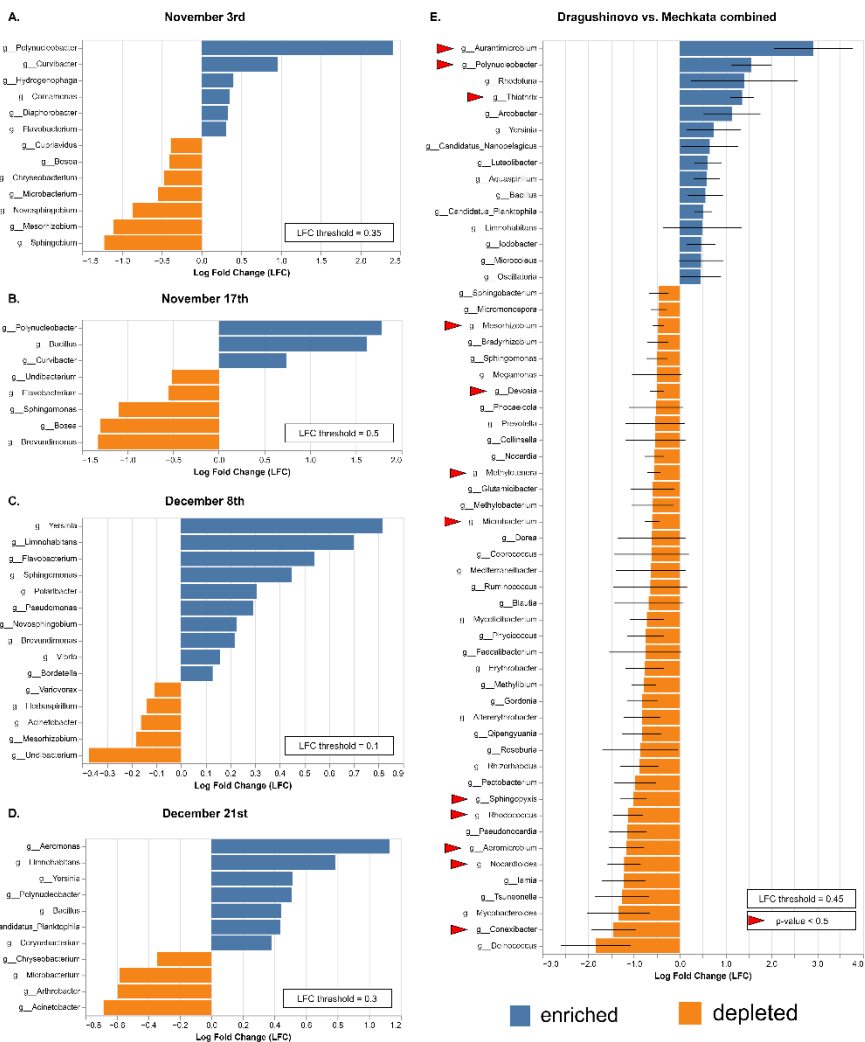
След това беше използвана PERMANOVA, за да се определи дали има значителни разлики в състава на бактериалните съобщества между пробите, събрани от различни места и дати, като бяха използвани индексите за бета разнообразие на Jaccard и Bray-Curtis (Виж *Фигура №10*). От припокриващите се таксономични профили по фактор местоположение (Виж *Фигура №10, горе*) стана ясно, че има висока степен на сходство в микробния състав между двете местоположения, но не е статистически значимо (Bray-Curtis: $F = 0.68$, $R^2 = 0.10$, $p < 0.50$; и Jaccard: $F = 0.72$, $R^2 = 0.10$, $p < 0.56$). От друга страна се наблюдаваше отчетливо разделение между пробите, групирани по дати (Bray-Curtis - $F = 4.89$, $R^2 = 0.79$, $p < 0.012$; и Jaccard - $F = 3.48$; $R^2 = 0.72$; $p < 0.007$), което показва, че таксономичните различия между датите са статистически значими и пробите се различаваха от дата до дата (Виж *Фигура №10, долу*).

Докато анализът на бета-разнообразието може да идентифицира прилики и разлики в таксономичния състав на съобществата, АДИ беше приложен, за да се определи кои таксони са движещи за тези различия, осигурявайки по-подробно разбиране за динамиката на промените в състава на бактериалните съобщества в р. Искър (Виж *Фигура №11*). Родът *Polynucleobacter* се класираше на 4-то място в общото разпределение (Виж *Фигура №8*), но беше определен като статистически обогатен ($p < 0.5$) в локацията надолу по течението и при четирите дати.



Фигура №10. Тестовите PERMANOVA, приложени към данните от Анализ на главните компоненти, използват индексите за разстояние на Bray-Curtis и Jaccard, за да се оценят разликите в състава на бактериалното съобщество между групите, като се обърне внимание на изобилието и наличието на видовете във всяка група. Резултатите не показаха статистически значими разлики между групите.

Polynucleobacter има много по-висока стойност ЛПП в ранните дати ($1.75 \div 2.4$) в сравнение с последните две дати ($0.4 - 0.5$), което може да се дължи на температурните вариации и кореспондира с предишни изследвания върху времевите промени в таксономията на сладководни екосистеми [25]. *Aurantimicrobium* присъстваше над 0.1% в 7/8 проби и беше значително увеличен ($p < 0.5$) във втората локация. Родът *Curvibacter* беше обогатен във втората локация през първите две дати на пробовземане. Наблюдавано беше изместване на диференциално обогатените бактерии между ранните и късните дати. На последните две дати бяха обогатени родове като *Yersinia* ($0.5 - 0.8$ ЛПП, $p < 0.5$), *Limnohabitans* ($0.7 - 0.8$ ЛПП, $p < 0.5$), *Flavobacterium* ($0.1 - 0.53$ ЛПП, $p > 0.5$), *Vibrio* ($0.09 - 0.15$ ЛПП, $p > 0.5$) и *Bordetella* ($0.04 - 0.13$ ЛПП, $p > 0.5$).



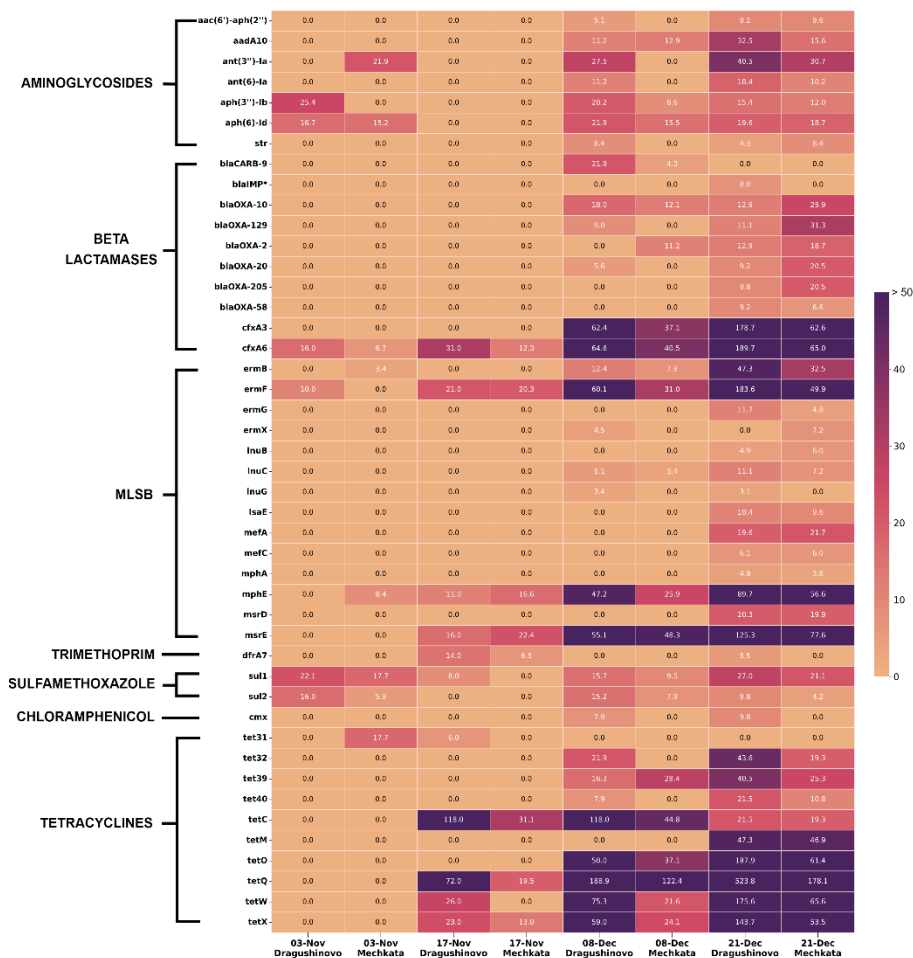
Фигура №11. Обогадени (сини) или намалени (оранжеви) таксони при АДИ на проби както пространствено, така и времево. Таксономичната таблица беше филтрирана, за да се запазят само таксоните, които присъстват във всички осем проби и съставляват $\geq 0.1\%$ от общото във всички проби. Пробите от локация с. Драгушиново (след ПСОВ) бяха съпоставени с пробите от Мечката (надолу по течението) за всяка от четирите дати на събиране: **A.** 3-ти ноември; **B.** 17-ти ноември; **C.** 8-ми ноември; **D.** 21-ви декември; **E.** Всички проби въз основа на местоположение. За **E.** бяха приложени по-отхлабени критерии за филтриране: включени таксони $\geq 0.1\%$ поне в две проби. Статистически значими ($p < 0.5$) резултати са посочени с червени стрелки.

4.4.2.3. ГАР и асоциация с потенциален гостоприемник

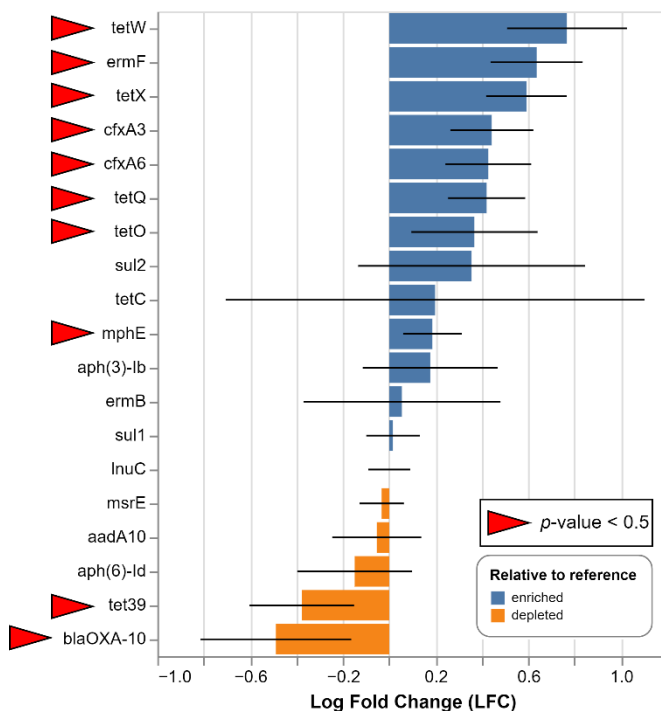
Първоначално бяха идентифицирани всички налични ГАР пробите посредством базата данни CARD. Тя открива максимално голям брой публикувани в литературата ГАР без значение техния произход. На жаргон много автори наричат този набор от наличните вродени и/или придобити ГАР фон (от англ. „background“). Не задължително този цялостен профил на ГАР е информативен за дискутиране. От значение е какви ГАР с клинично значение се откриват измежду тоталната извадка. Поради това клинично значимите ГАР трябваше да бъдат отсети посредством други приложения, които разполагат с БД само с ГАР, за които има литературни данни, че в изолирани условия доказано водят до повишаване на минимална потискаща концентрация (МПК). Тези ГАР са обобщени на *Фигура №12*. Сред тях най-многобройни са групите, придаващи резистентност към макролиди и тетрацилини, следвани от бета-лактамите, аминогликозидите (главно стрептомицин) и сулфонамидите. В пробите от последните две дати и по-специално в с. Драгушиново, разположена непосредствено след ПСОВ гр. Самоков, беше наблюдавана ясна тенденция на увеличаване на относителното количество ГАР. Беше установено, че редица ГАР са значително обогатени в участъка в с. Драгушиново (Виж *Фигура №12* и *Фигура №13*).

Този резултат не може да бъде дело на разлики в дълбочината на секвениране, тъй като броят на прочитите от данните, получени от всички проби, бяха съпоставими, а в първите четири проби някои от тях напълно липсваха. Също така, същият анализ беше повторен с извадки от случайни прочити от последните четири проби и моделът на heatmap графиката остана подобен. Интересно беше откриването на две карбапенемази - *blaOXA-58* и *blaIMP*-подобен.

Генът *blaOXA-58* беше открит с 97.51% идентичност и покритие и също така идентифициран в ко-асембли C2-Dragushinovo със 100% съвпадение. По отношение на *blaIMP* няколко прочита от S7-Dragushinovo частично се залепиха върху гена *blaIMP-13* (77.46% идентичност и покритие), докато генът *blaIMP-33* със 100% покритие и 95.5% идентичност се появи в рамките на ко-асембли C1-all. Необходими са допълнителни изследвания, за да се установи точния вариант, но съществува вероятност за наличието на няколко варианта на IMP въз основа на това, как всички прочити се наслаждава към референтните секвенции на IMP-13 и IMP-33. Въпреки това самото присъствие на двата гена в сладководната река Искър е тревожно, тъй като те обикновено се откриват в болнични условия и понякога в животновъдни ферми.



Фигура №12. Heatmap графика от нормализираните прочити, които успешно се налагаха върху GAT с клинично значение. Включени бяха само тези, детектирани поне в две проби, с изключение на гена *blaIMP*, който е отбелязан с *. Отделно всички бяха открити с минимум 60% покритие и 60% идентичност. Гените са подредени въз основа на класа антимикробни средства, към които придават резистентност. Всички гени с $n=50+$ прочита са оцветени в един и същи цвят с цел визуализация. MLSB се отнася до GAT, които придават резистентност към един или няколко от следните антибиотици: еритромицин, азитромицин, телитромицин, линкозамид, стрептограмин В/А, линкомицин, клиндамицин, хинупристин и далфопристин.



Фигура №13. АДИ беше използван за идентифициране на обогатени (синьо) или ниско представени (оранжево) ГАР въз основа на техните стойности ЛПП в първата локация с. Драгушиново (по-близо до ПСОВ Самоков). С течение на реката, тези ГАР намаляваха. Включени са само такива, открити в поне 4 проби. Статистически значимите ($p < 0.5$) резултати са обозначени с червени стрелки.

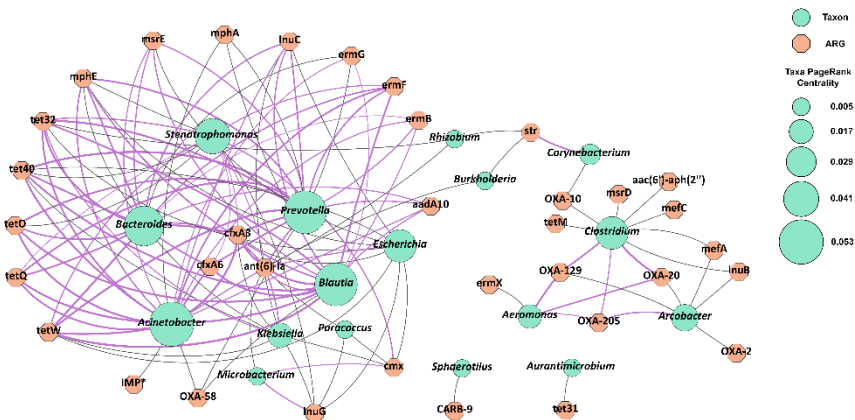
За целите на намирането на предполагаем гостоприемник на всички клинично значими ГАР във вид на функционална таблица бяха извлечени брой сурови прочити, които отговарят с висока степен на сигурност на тези ГАР. Функционалната и таксономичната таблица заедно бяха използвани за изчисляване корелацията на Pearson между тях и откритите таксони (Виж *Фигура №14*). Оформиха се два клъстера, като единият от тях включваше многобройни ГАР за тетрациклини (*tetQ*, *tetW*, *tetO*, *tet40*, *tet32*), макролиди (*ermB*, *ermF*, *ermG*, *mphE*, *msrE*, *mph*), и бета-лактамазни гени и ключовите таксони *Acinetobacter* (18 ребра) *Bacteroides* (15 ребра), *Prevotella* (17 ребра), *Blautia* (17 ребра), *Stenothrophomonas* (12 ребра), *Escherichia* (10 ребра), *Klebsiella* (9 ребра). Голяма част от тези ГАР като *tetQ*, *tetW*, *tetO*, *ermF*, *mphE*, *cfxA3*, *cfxA6* и *ermB* бяха също така намаляваха ($p < 0.05$) с течението на реката към втората локация както е показано на *Фигура №12* и *Фигура №13*. Успоредно с тях родовете *Prevotella*, *Blautia*, *Bacteroides*, *Escherichia*, *Klebsiella*, *Acinetobacter* и *Clostridioides*, с които те корелираха положително ($p < 0.05$), също значително намаляваха от първата

към втората локация. Наблюденията, че клинични значими ГАР и представители на патогенни за човека бактерии бяха с увеличена релативна честота при заустването на рециклираните води от ПСОВ и заедно намаляваха сигнификантно с течението на реката, беше индикаторно, че те не са естествено разпространени във водата и не са част от нейния микробиом. Най-вероятно те са били внесени със заустването на водите от ПСОВ гр. Самоков и с течение на реката те се разреждат. Забележително беше, че сигнификантни положителни корелации бяха открити само за тези микроорганизми измежду огромния набор от бактериални родове в таксономичния състав и само за фракция от всички ГАР. Също така, голяма част от тях са или опортюнистични човешки патогени или са характерни за човешките микробиоми, а не за микрофлора на сладководен басейн.

От друга страна, двата рода *Prevotella* и *Blautia* бяха открити сред първите 30 най-представени рода в пробите, а те са едни от най-често използваните индикатори за фекално замърсяване на води от хора. Установено е, че и двете бактерии намаляват (несигнификантно) във всички двойки проби между двете местоположения, а също така положително корелират със едни и същи 17 клинично важни ГАР за макролиди и тетрацилини. Други проучвания също откриват ГАР като *cfxA3*, *tetQ* и *ermF* в *Prevotella*, като *tetQ* често се намира върху конюгативен транспозон или заедно с *ermF*. Освен това в ПСОВ Самоков липсва третично пречистване, за което е доказано, че допълнително намалява в пъти бактериалната концентрация в отпадните води.

Също така ГАР, открити във водите, бяха в голяма степен съвместими с резултатите от други проучвания на сладководни водни системи: такива, които придават резистентност към аминогликозиди, като например аминогликозидните фосфотрансфери (3')-*la*, (3')-*lb* и (6)-*ld*; нуклеотидилтрансфери, често срещани в активни утайки на ПСОВ *aadA*, *aacA4*, *aadB*, *aadE*, *strB*; такива, които придават резистентност към макролиди, като *ermB*; *msrE* и *mphE*; *sul1* и *sul2*; и *tet* гени. Най-обезпокоително беше ясното присъствие на карбапенемазните гени *blaOXA-58* и *blaIMP-33*-подобен. Генът *blaOXA-58* често се среща в изолати на *Acinetobacter* в клинични условия и е най-разпространеният в няколко страни в Европа. Най-важното е, че той придава резистентност към антибиотиците за лечение от последна линия - имипенем и меропенем. Ензимът преобладава и във води от крайбрежните райони и е свързан с различни бактерии-гостоприемници. Освен това е установено, че абсолютното количество на гена *blaOXA-58* е по-високо в отпадъчните води от ПСОВ в сравнение с местата нагоре или надолу по течението спрямо други изследвани ПСОВ, което съответства на настоящите резултати. В тази връзка сравнението на придобитата резистентност и съвместната ко-локализация на ГАР и ПГЕ установи повече попадения в първото местоположение Драгушиново, т.е. непосредствено след ПСОВ, като по този начин се съгласува с резултатите от други проучвания. Също така се наблюдава повишаване на изобилието на резистома и мобилома на рециклираните води от ПСОВ в сравнение с местата нагоре или по-надолу по течението на реките, в които се заустват. Тези доказателства допълнително засилват нуждата от откриването

и докладването на ГАР в околната среда и допринасят за разбирането, че отпадъчните води от ПСОВ могат да служат като движещ фактор за оформяне на водите от приемащия воден басейн.



Фигура №14. Двустранна мрежова графика, показваща асоциациите гостоприемник-ГАР. Включени са само най-силно представените 40 таксона и конкретни известни човешки патогени. ГАР, открити само в по-малко от две проби, бяха изключени, с изключение на гена IMP, който е отбелязан с *. Зелените кръгли възли (от англ. “Node”) представят бактерии, докато оранжевите осмоъгълници са ГАР. Ребрата (от англ. “Edge”) свързват възли от двете групи въз основа на положителна Pearson корелация (≥ 0.8) и р-стойност (≤ 0.005), а цветът и дебелината им означават тежестта на връзката (лилавото е по-високо). Положителните корелации показват висока теоретична вероятност даден таксон да е гостоприемник за ГАР. Размерите на възлите на родовете означават централността на всеки възел в мрежата. Визуалната дължина на ребрата трябва да се пренебрегне, тъй като оформлението на графиката е коригирано ръчно, за да се подобри четливостта. Не всички връзки в мрежата могат да представляват значими и истински взаимоотношения.

През 2001 г. ПСОВ Самоков е най-модерната пречиствателна станция на Балканите с механично, биологично управление и управление на активните утайки, но без третично третиране. Поради своя огромен капацитет от 125,000 еквивалент жители, ПСОВ гр. Самоков би трябвало да бъде включена в редовното наблюдение на АМР върху отпадъчните води и ПСОВ, както е видно от предложената директива от Европейската комисия.

4.4.2.4. Съвместна поява на ГАР и ПГЕ в предполагаеми плазмидни последователности и оценка на риска на резистома

За разлика от резултатите посочени до момента в пробите от р. Искър, за които бяха използвани само суровите прочити, за идентификация на плазмиди и откриването на ко-локализиран ГАР и ПГЕ върху плазмиди, е нужно да се извърши

асемблиране. Поради недостатъчна дълбочина на секвениране и генерално ниската концентрация на ГАР беше избран подходът за ко-асемблиране, което е обединяване на всички сурови данни в един общ пул и общото им ко-асемблиране. Бяха създадени три набора от сурови прочити чрез обединяване на всички необработени прочити (C1-all) и по местоположение (C2-Dragushinovo, C3-Mechkata).

Трите ко-асемблирани метагенома (C1-all, C2-Dragushinovo и C3-Mechkata) бяха подложени на три различни инструмента за прогнозиране на плазмиди. Резултатите от инструментите имаха различни степени на съответствие, като някои контиги бяха идентифицирани като плазмидни от няколко инструмента, а повечето - само от един инструмент. За да получим цялостна представа за потенциалните плазмиди, включихме всички контиги, маркирани като плазмидни от поне един инструмент. В резултат в ко-асемблитата C2 и C3 съответно 3386 и 1897 контига бяха маркирани като плазмидни. Те бяха отделени и върху тях бяха приложени апликациите за откриване на ГАР и ПГЕ. В резултат, бяха идентифицирани общо три случая на ко-локализация на ГАР и ПГЕ върху плазмиди: *blaOXA-58* и *ISAlw25*, открити само в C2-Dragushinovo; *Tn917* и *ermB* само в C2-Dragushinovo; и *Tn6082*, носещ *aph(6)-Id* и *aph(3'')-Ib* в C2 и C3. Интересно е, че по-голям брой предполагаеми плазмиди и повечето случаи на съвместна поява бяха открити в локацията с. Драгушиново, която е веднага след точката на заустване, въпреки че асемблитата имат сравними N50 (2159 нб срещу 2196 нб) и обща нуклеотидна дължина (621M нб срещу 475M нб).

Карбапенемазата *blaOXA-58* и *ISAlw25* бяха разположени на контиг от 1256 нб с 34 нб разстояние между тях, което на свой ред беше идентифицирано от PlaSquid и PLASMe. В генбанките най-близкородственият плазмид на него беше *pOXA58_005069* (NZ_CP026086.2) от *Acinetobacter pittii*, докато според ISFinder *ISAlw25* е открит в плазмид на *A. lwoffii*. Суровите прочити, отговарящи на *blaOXA-58*, бяха открити само при последната дата на вземане на проби (Виж *Фигура №12*). Напълно реконструиран ген със 100% покритие и идентичност беше постигнат само в ко-асемблитата C1-all и C2-Dragushinovo, но не и от C3-Mechkata, което показва по-голямото изобилие на гени *OXA-58* в точката на заустване. *Tn917* (*Tn551* в *TnCentral*), носещ *ermB*, е разположен върху плазмиден контиг от 5717 нб, идентифициран от PlaSquid и PLASMe. Генът за макролидна резистентност *ermB* е нативен елемент на *Tn917* и също така е широко разпространен в *Streptococci*, *Enterococci* и *Bacteroides*, което се съгласува добре с анализа на асоциацията на *ermB* с *Bacteroides*, *Prevotella*, *Acinetobacter* и *Blautia* (Виж *Фигура №14*). *Tn6082*, носещ *aph(6)-Id* и *aph(3'')-Ib*, е широко разпространен в клинични щамове на *Enterobacterales* и *Pseudomonas*. Той е открит в изолати *E. coli*, продуциращи бета-лактамаза с разширен спектър (БЛРС) от прасета в Южна Африка и в изолати *Klebsiella* от сладководно езеро в Индия.

Плазмидите също се разглеждат като ПГЕ и въпреки, че съвместната поява на ПГЕ и ГАР върху плазмиди може да е ограничена поради фактори като непълно

асемблиране или дълбочина на секвениране, е важно да се отбележи наличието на останалите ГАР, идентифицирани само върху предполагаеми плазмидни секвенции, поради техния потенциал за хоризонтален генетичен пренос (ХГП). На двете локации са открити следните допълнителни гени: *aph(2'')-I_f* (амикацин, гентамицин, тобрамицин, канамицин); *blaOXA-1* (ESBL), *blaOXA-10* (ESBL, aztreonam) и *blaOXA-20* (пеницилини); *sul1* и *sul2* (сулфаметоксазол); *tet(39)*, *tet(C)*, *tet(O)*, *tet(W)* и *tet(X2)*, които заедно придават резистентност към доксициклин, тетрациклин, миноциклин и тигециклин; *erm(B)*, *mph(E)*, *msr(E)*, които заедно придават резистентност към еритромицин, азитромицин, клиндамицин и др. Последните три бяха локализирани върху един контиг (в C2-Dragushinovo) с по-високо секвенционно покритие в сравнение с други ГАР, което подчертава тяхното изобилие и предполагаемо съвместно предаване.

Беше използван софтуер MetaCompare 2.0 за изчисляване на оценката на риска от резистомии за двете локации и за реката като цяло. Този софтуер на практика извършва описаните вече от нас анализи, но също така търси ГАР върху контиги с хромозомен произход и идентифицира патогени. Комбинираният резистомен рисков фактор беше определен само върху трите ко-асемблирани метагеноми. Резултатите от C1-all бяха разглеждани като представителна оценка за цялото изследване върху р. Искър, в този регион и в този период (Виж Таблица №13). Рисковият фактор възлизаше на стойности 6.28 за екологичните ниши и 0.19 за човешкото здраве, което се счита за нисък риск според MetaCompare 2.0 [15]. Това беше в съответствие с малкия брой открити съвместни случаи на ГАР и ПГЕ. Оценките на риска по местонахождение (C2 и C3) бяха много сходни с тоталната оценка.

Таблица №13. Резултати от MetaCompare 2.0. Най-отдолу са стойностите на рисковият фактор.

	C1-all	C2-Drag.	C3-Mech.	C1-all	C2-Drag.	C3-Mech.
score type	Ecological			Human health		
Risk Score	6.2	6.3	5.7	0.18	0.18	0.16

4.4.2.5. Реконструирание на МАГ

От ко-асембли C1-all бяха получени общо 29 МАГ. От трите приложения за биниране бяха създадени бинове с $\geq 70\%$ завършеност и $\leq 10\%$ замърсяване, след което бяха консолидирани и усъвършенствани със софтуер за бинове на MetaWrap. Консолидирането сравнява получените бинове от трите приложения и ако един и същи бин е генериран от трите приложения, то той се отделя като надежден. След това консолидираните бинове се подлагат на повторно асемблиране (още ре-асемблиране) за подобряване на качеството на геномите.

Получените МАГ бяха оценени по отношение на качеството с Busco и CheckM и подадени за обществено ползване към генбанката NCBI (PRJNA1071831). От всички тях пет бяха висококачествени ($\geq 90\%$ завършеност и $\leq 5\%$ замърсяване), а останалите бяха със средно качество според публикувани критерии. Всеки МАГ беше

идентифициран до възможно най-ниския таксономичен ранг. Само два от тях носеха ГАР. Това бяха (bin.2.orig), класифициран в род JAAFJM01 (*Bacteroidota bacterium*), носеше аминокликозид модифициращи *aph(3'')-Ib* и *aph(6)-Id*, и род *Aurantimicrobium* (bin.19.orig) носеше ген *tet* с висока идентичност с *tet(37)*. За друг МАГ, *Agathobacter rectalis* (bin.12.permissive), беше установено, че носи *ermG*, а бактерията е естествено открита в трбуха на крави и овце, което показва потенциално замърсяване с животински фекалии.

Това е едно от малкото проучвания за България, които успешно са реконструирали МАГ от водни проби. Едно подобно проучване с реконструиране на МАГ е проведено до сега върху термофилни микроорганизми от български горещи извори [26].

4.4.2.6. Недостатъци на проучването и заключение

Могат да бъдат отбележат няколко недостатъка на проучването. Пробите от първата локация не са преки представители на отпадъчните води от ПСОВ Самоков. Първата локация с. Драгушиново е на около 800 м надолу по течението от действителното място на заустване и не беше възможно оценяването на абсолютния мащаб на въздействие на отпадъчните води върху мобилома и резистома на реката. Бъдещите ни проучвания могат да бъдат потенциално подобрени чрез увеличаване на локациите и прецизиране на местата за пробовземане, като напр. локации нагоре по течението или изходяща вода от ПСОВ, за да се опише по-точно въздействието на ПСОВ върху резистома на реката. КНИ на приложения МТС като контрола на екстракцията възлизаше на 67 точки, което означава, че се наблюдават нива на отклонение, внесени в настоящия анализ от изолирането на ДНК и секвенирането. Това предполага допълнително оптимизиране или по-внимателна работа по време на обработването на пробите.

В заключение, това е първото всеобхватно метагеномно проучване, насочено към документиране на бактериалното таксономично разнообразие на сладководна река в България, и ефектите ПСОВ гр. Самоков върху резистома и мобилома. Към момента на изготвяне на доклада в България има само едно публикувано проучване от 2017 г. за 16S рРНК метагеномно профилиране на бактериални сладководни съобщества в два български язовира [27], което показва голямата неизследвана област. По отношение на АМР могат да се изброят няколко опита за откриване на АМР във водни обекти, но всички те са базирани на културелни методи [28]. Също така, има едно молекулярно-базирано проучване за целенасочено откриване на ГАР в язовир Искър, който се намира надолу по течението на реката от нашите две локации. Обаче, то е с фокус върху нетуберкулозните микобактерии, пренасяни по вода [29].

4.4.3. Анализи, базирани на ПСОВ гр. Стара Загора и ПСОВ птицекланица „Евровърхът“

4.4.3.1. Устройство на ПСОВ

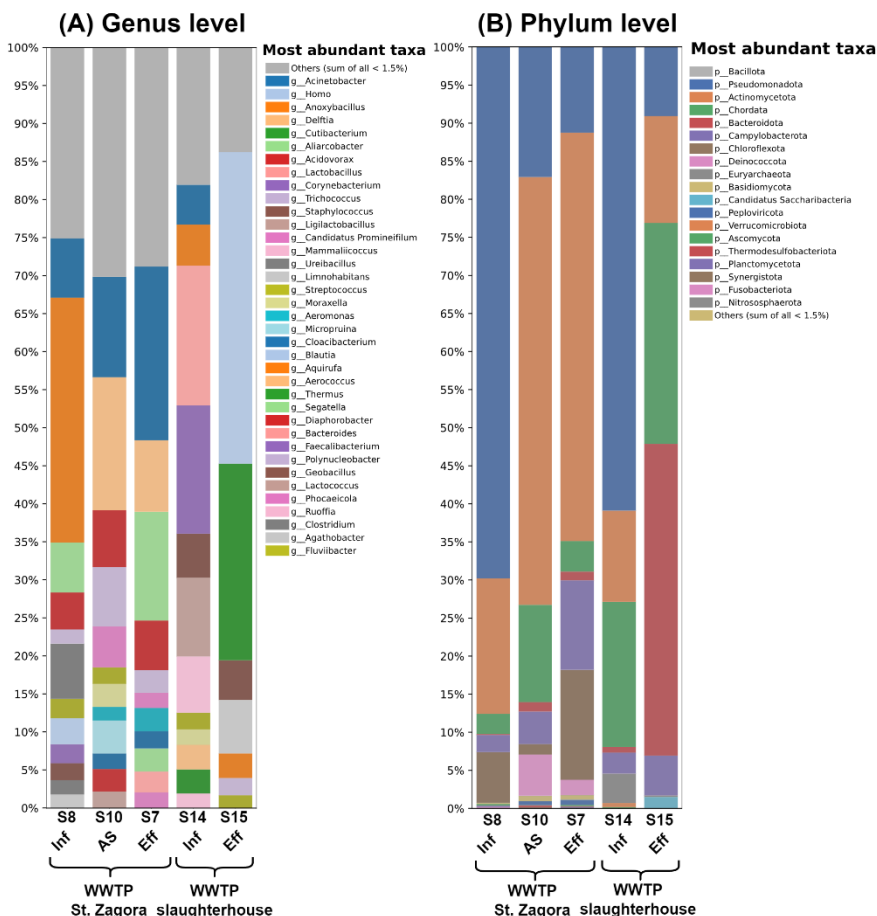
ПСОВ гр. Стара Загора е в експлоатация от 2011 г. и е собственост на Община Стара Загора. Проектирана за 256,300 еквивалент жители и $Q_{ор}=6827 \text{ м}^3/\text{ч}$, биологично натоварване на кислород от 13,841 кг/д и суспендирани вещества натоварване от 15,954 кг/д. По информация на „ВИК“ Стара Загора тя е проектирана за пречистване на отпадните води от града при смесена канализационна система като разполага с механично и биологично пречистване на принципа на активната утайка, третиране на утайките в анаеробен изгнивател, дезинфекция с хлориране, включваща контактен резервоар и хлораторна инсталация, и ко-генерация за добиване на ел. енергия от биогаза. Според тази информация, ПСОВ гр. Стара Загора може да се класифицира като второ ниво на пречистване (механично и биологично) с компоненти от трето ниво на пречистване (дезинфекция чрез хлориране).

Локално разположената ПСОВ в птицекланица „Евровърхът“ обслужва само своите нужни и работи само по време на колежа на птици. Тя разполага само с първично и вторично пречистване без стъпки на третично пречистване.

4.4.3.2. Таксономичен състав

На *Фигура №15* са представени най-широко разпространените бактериални представители по тип и род. Един от най-широко представените таксони беше род *Anoxybacillus* с малко над 40% в мръсните отпадни води, вливащи се в ПСОВ гр. Стара Загора. А от тях над 26% се падат на вида *Anoxybacillus flavithermus*, който е известен с това, че образува биофилми в млекопреработвателната индустрия. *Anoxybacillus flavithermus* е сравнително нов вид, регистриран през 2000 г. Според други автори, заедно с *Geobacillus* spp. те са основни замърсители в млечната промишленост. *Geobacillus* spp. беше регистриран в малко над 2% в S8. Също така, двата вида се откриват в сухо мляко. Те са считани за добри агенти за биопроцесиране и биотрансформация, което ги прави привлекателни за индустрията. Други микроорганизми с висок процент на изолиране, за който също е известно, че най-вероятно произхождат от хранителната промишленост, са *Anoxybacillus* spp. (> 14% в S8), *Ureibacillus thermophilus* (8.3% в S8), за които е характерно, че се изолират от компост от животински тор. Още, *Acinetobacter johnsonni* (4.5%), който е характерен за отпадъчни води и *Blautia wexlerae* (3.3%), който са характерен както за човешки така и за животински фекалии, бяха открити в отпадъчните води на гр. Стара Загора (S8). В заключение, таксономичното разнообразие на отпадъчните води съответства на други отпадъчни води и е съставено от микроорганизми, които са фекални замърсители, част от човешката чревна микрофлора или типични за селското стопанство или млекопроизводство.

Проба S10 е активна утайка. Генерирането на активна утайка е най-широко застъпеният процес по време на вторичното пречистване на водата. Той е вид



Фигура №15. Барплот графики с таксономичен състав на ниво род (А) и тип (В), изобразяващи относителното изобилие на най-разпространените таксони във всяка от пробите, подредени по източник и по посока на течение на водата от входяща (Inf) през активна утайка (AS, само за ПСОВ Ст. Загора) към изходяща (Eff). Нискоразпространените таксони (< 0.1 %) и таксоните, открити в по-малко от 2 проби, бяха отстранени. Таксони с изобилие под 1.5% бяха групирани в група "Others". Най-широко разпространението таксони са оцветени в различни цветове.

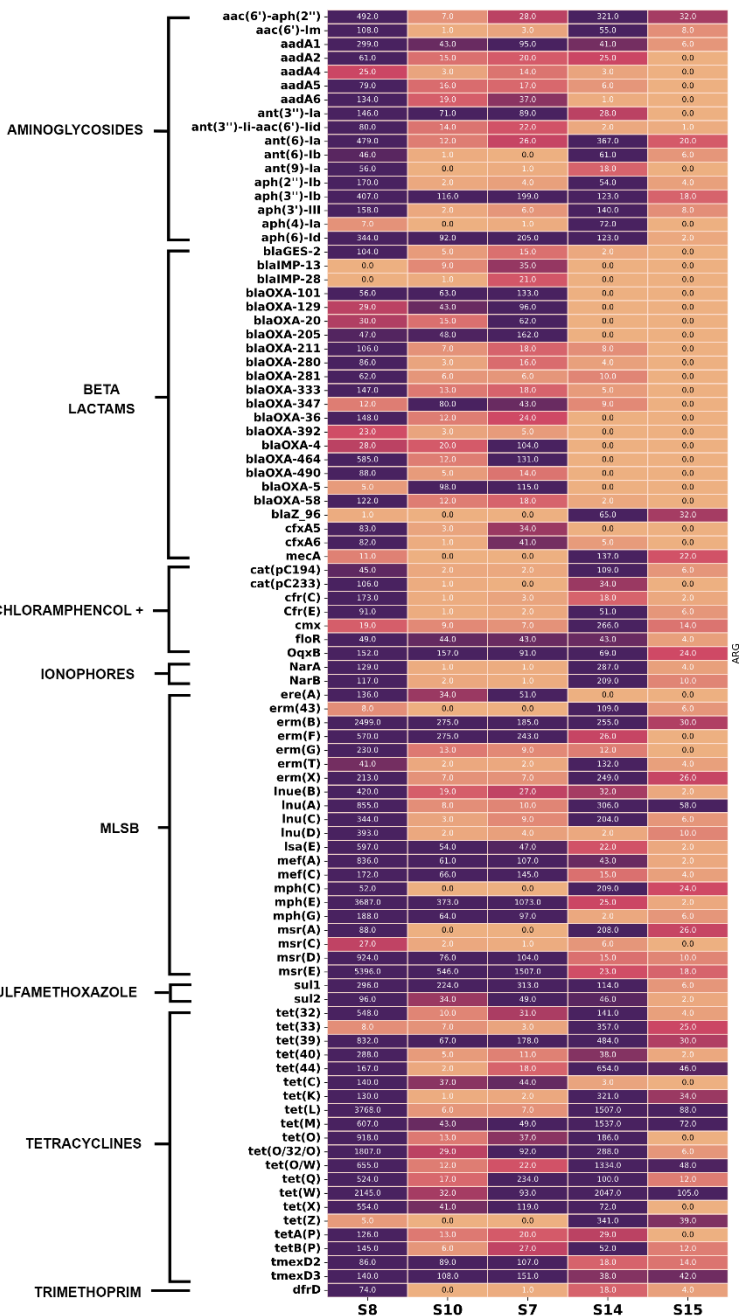
биологичен процес на третиране на канализационни или промишлени отпадъчни води. При него се използва биологична маса, съставена от бактерии и протозои, които веднъж добавени в специални резервоари наречени биобасейни на станцията се мултиплицират и изграждат локална микрофлора и биофилм. Този процес се активира като средата в биобасейните се аерира. Микроорганизмите се размножават, разграждат органични вещества от мръсната вода, продуцират крайни и междинни

метаболити като в резултат се образува отпадъчна утайка (или флоки). Водата, преминала през аериращия резервоар, се прехвърля в утайтелен резервоар, който има конусовидно дъно. Тя престоюва в покой докато формираните флоки се утаят на дъното. От там водата над утайката се прехвърля към следващи стъпки на пречистване, които са хлориране в настоящия случай. От друга страна утайката се неутрализира, съгъства и изсушава. В ПСОВ гр. Стара Загора част от нея се подлага на процесът за генериране на биогаз.

В S10 бяха идентифицирани типично характерни за активна утайка микроорганизми като *Acidovorax temperans* (> 8.2%) и *Trichococcus flocculiformis* (> 9.0%), а също и *Acinetobacter johnsonii* (>8.3%). Микрофлората на рециклираната вода до голяма степен наподобяваше състава на активната утайка, но в значително променено разпределение. Например, род *Aliarcobacter* от 1.8% се е увеличил до 19.7% в рециклираните води, последван от *Acinetobacter*, който се увеличава с близо 7% до 21.76%. От друга страна, процентната честота намалява на представители като род *Trichococcus* (повече от 2 пъти) и род *Delftia* (близо 2 пъти). Може да се каже, че хлорирането променя профила на микробиома, но качествено той остава почти същият. Количествен анализ е нужен, за да се определи с каква скорост намалява тоталното количество на микроорганизми. От *Фигура №15В* може да се види тенденция на изместване на тип *Pseudomonodota* с *Actinomycetota* вследствие от пречистването на водата. Този тип промяна е характерна за ПСОВ от второ ниво на пречистване, дори и с третични стъпки като хлорирането. Според автори, измервали съдържанието на общ органичен въглерод, общ фосфор и консумация на кислород в изследваните проби, наличието на *Actinomycetota* във високи концентрации в активните утайките има отрицателно въздействие върху качеството на приемника на отпадъчни води. Също така, тяхната относително висока релативна честота влияе неблагоприятно върху ефективността на процеса на пречистване на отпадъчните води. Въпреки това те са ефикасни в биоразграждането на пестициди, както и в разграждането на биомолекули, намалявайки стойностите на биохимичната и химичната потребност от кислород и тотално суспендирано вещество, както и да намаляват концентрациите на тежки метали като Cu, Fe, Mn, Pb и Zn. Също така, редица *Actinomycetes* са продуценти на биоактивни съединения с антибактериална активност и потискат разпространението и размножаването на Грам-позитивни и Грам-негативни микроорганизми сред които и патогени като *Staphylococcus aureus* и по-малко *Pseudomonas aeruginosa*.

4.4.3.3. Детектиране на ГАР

На *Фигура №16* са представена откритите ГАР. За целта отново бяха приложени 60%/60% идентификация и покритие. Приложенията като Resfinder са изключително полезни за този тип анализ, но те имат едно несъвършенство. Всяка референтна последователност на ГАР в тяхната база данни се анализира поотделно и поради това суровите прочити могат да се наложат върху повече от една референтна



Фигура №16. Heatmap графика, показваща нормализираните прочити, които успешно се налагаха върху ГАР с клинично значение. Включени бяха само ГАР с повече от 1.5% брой прочити, спрямо тоталното за цялата проба, с изключение на някои гени, които се срещат често по литературни данни и могат да се пренасят с ПГЕ. Отделно, всички включени ГАР, бяха открити с минимум 60% покритие и 60% идентичност. Всички гени с $n=50+$ прочита са оцветени в един и същи цвят с цел визуализация. MLSB се отнася до ГАР, които придават резистентност към един или няколко от следните антибиотици: еритромицин, азитромицин, телитромицин, линкозамид, стрептограмин В/А, линкомицин, клиндамицин, хинупристин и далфопристин. Групата Chloramphenicol + включва резистентност към: амфеникол, хлорамфеникол, оксазолидинон, линкозамид, стрептограмин А, флорфеникол, клиндамицин, линезолид, ципрофлоксацин, триметоприм.

секвенция от БД. Съответно, алели на ГАР като *tet*, *mph*, *msr* или *blaOXA* могат да бъдат детектирани с голям брой прочити, но в действителност повечето от тези прочити произлизат от значително по-малко разнообразие на ГАР в пробата. Това разбира се, беше необходим риск, за да не бъдат изпуснати гени. От друга страна, повторение на анализа с прилагане на 100% покритие и идентичност, гарантирано показва, че те са налични в пробата. Всички дискутирани ГАР по-долу са със 100% показатели, а графиката тип heatmap (Фигура №16) е с цел по-добра визуализация на целия спектър от потенциални ГАР.

В проба S8 (канални отпадъчни води, вливащи се в ПСОВ гр. Ст. Загора) със 100% идентичност, 100% покритие и със 100% дълбочина на покритието бяха детектирани следните ГАР:

- Гена *blaOXA-1* (БЛРС), разпространена в *Enterobacteriales*, често откривана в *E. coli* и *Klebsiella pneumoniae* заедно с други БЛРС като CTX-M-15;
- Гена *blaOXA-10* (БЛРС) продуцира високо стабилна бета-лактамаза. Генът е с широк спектър на гостоприемници сред човешки нозокомиални Грам-негативни патогени. Това предимно са *Enterobacteriales*, но също така и в *Pseudomonas*. Генът се открива в интегрон клас върху IncY конюгативен плазмид. Това показва наличието на движещи фактори на ХГП;
- Карбапенемазата *blaOXA-58* беше открита в пробата от мръсните води преди ПСОВ Ст. Загора, както и в рециклираните води. Наличието и преди и след ПСОВ кореспондираше с находките от проучванията в р. Искър. Този ген явно е често срещан за канални води и със сигурност се изпуска посредством рециклираните води навън в природата;
- Друг ген беше *blaOXA-101* (сем. OXA-10) с тесен спектър на бета-лактамна хидролиза. Интересното при него е неговата локализация в клас 1 интегрони върху конюгативен плазмид при *Enterobacteriales*;
- БЛРС *blaOXA-129*. Този ген се среща в *P. aeruginosa*. Още, генът *blaOXA-333* открит хромозомно разположен в *Acinetobacter johnsonii*. Няма литературни данни за плазмидно пренасяне и най-вероятно има много тесен спектър на

действие, който не е напълно проучен. Наличие на този ген не е случайност, защото *Acinetobacter johnsonii* се открива в трите проби от ПСОВ гр. Стара Загора (4.5% в S8, 8.3% в S10 и 9.3% в S7), което го прави сравнително постоянен и равномерно представен по целия път на рециклиране на отпадъчните води;

- Генът *bla*ACI-1 е слабо познат в клиничната микробиология. Има данни за неговото детектиране в човешки черва в представители на *Negativicutes*, но е силно разпространен в микробиома на току що отбити прасета. *Negativicutes* не бяха детектирани, но голям процент се пада на таксони, характерни за чревните микробиоми на селскостопански животни;
- Генът *bla*GES-5 е широко разпространен БЛРС в световен мащаб ГАР, кодиращ бета-лактамаза от Ambler клас A, за които автори пишат, че има карбапенемазна активност. Първоначално е открит в клиничен изолат на *E. coli*, след това в по-широк спектър от гостоприемници като *K. pneumoniae*, в *E. cloacae* и *P. aeruginosa*. Генът винаги е разположен на плазмид и е част от клас 1 интегрони. Често бива открит в ПСОВ или канални води.

Сред ГАР, водещи до резистентност към макролиди, най-много сурови прочити бяха регистрирани за *ermB*, *ermF*, *mefA*, *mefC*, *IsaE*, *mphE*, *mphG*, *msrE* и *msrD*. Всички останали варианти в групата макролид, линкозамид, стрептограмин Б (МЛСБ) бяха представени с много по-малко покритие и голямата част от тях се срещаша само в проба S8 (канални води) или драстично намаляваха вследствие от рециклирането. ГАР срещу сулфонамиди *sul1* и в по-малък аспект *sul2* се срещаша в почти непроменени относителни количества между трите проби, което може да бъде интерпретирано, че те са депозиранни в микроорганизми със сравнително постоянна релативна честота.

За разлика от водите от р. Искър, в каналните водите от гр. Стара Загора и във входящите мръсни води на ПСОВ в птицекланица „Евровърхът“ (S14) бяха открити два гени *narA* и *narB*. През 2016 г. за тях е доказано, че се кодират в оперон и заедно сформират АВС-тип транспортер, който води до резистентност към йонофорите наразин, салиномицин и мадурамицин. В цял свят наразин и салиномицин са молекули, включени като фуражни хранителни добавки в храните при отглеждане на домашни птици за борба с кокцидиозата. В допълнение, те притежават и антибактериална активност предимно срещу Грам-положителни бактерии. Техният механизъм за резистентност е детектиран за първи път в *Enterococcus*. Резистентни към наразин бактерии не представляват клиничен проблем в човешкото здравеопазване. Авторы, обаче, установяват, че резистентността към наразин и резистентността към ванкомицин могат да бъдат ко-локализиранни върху едни и същи плаزمиди. Това крие опасност от ко-селекция на резистентност към ванкомицин още на ниво отглеждане на домашни птици посредством йонофори в хранителните смеси.

Хлорирането като третична стъпка на пречистване на отпадъчните води има вариращ ефект спрямо ГАР и БРА. Проучванията показват, че хлорирането не намалява значително ГАР и дори в някои случаи повишава релативната честота на БРА [30]. Различията са основани на различни режими на хлориране (концентрация и време на експозиция). По литературни данни в диапазона между от 5 до 9 мг/л за 30 мин се наблюдава преференциално увеличение на ГАР срещу макролиди, тетрациклини, сулфонамиди, и по-рядко бета-лактами, аминогликозиди, рифампицин и ванкомицин. Интересни са наблюденията, че докато хлорирането намалява значително нивата на БРА като цяло, не намалява значително нивата на ГАР.

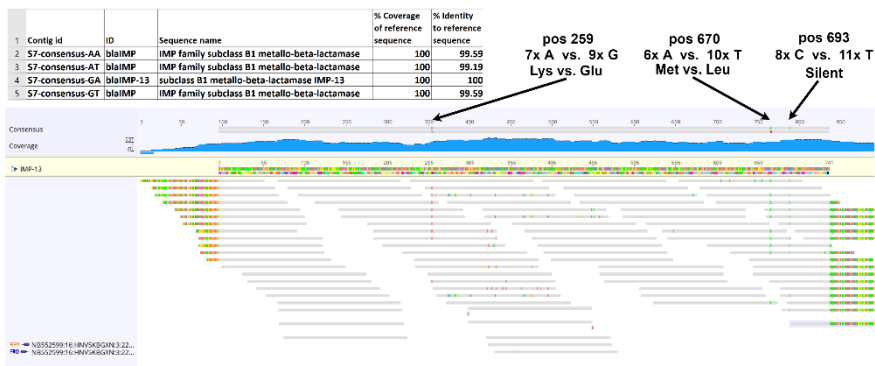
4.4.3.4. Продуцент на гена *blaIMP* в ПСОВ гр. Стара Загора

В асемблираните метагеноми бяха изследвани ГАР чрез приложението AMRFinderPlus, което работи единствено върху асемблирани геноми. Негово предимството е, че успоредно с нуклеотидната последователност извършва идентификацията и чрез протеиновата последователност. По този начин тихи мутации може да бъдат пренебрегнати. За разлика от пробите от р. Искър, дълбочината на секвениране на настоящите проби беше по-висока и наличието на ГАР много по-изразено, защото рециклираните води не са премесени в реката, в която се заустват. В резултат, асемблираните метагеноми на отделните проби са много по-богати на ГАР в сравнение с пробите от р. Искър, за които се наложи да се използва подхода ко-асемблиране.

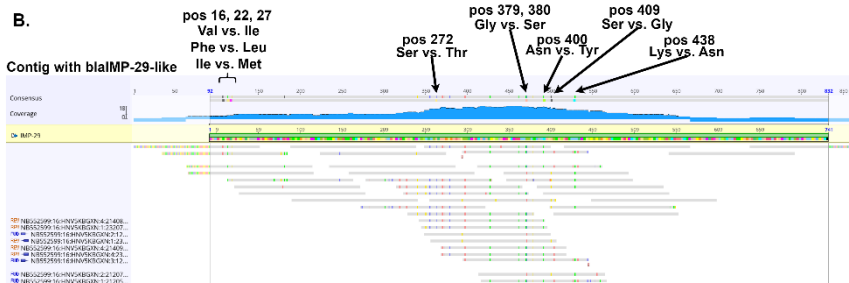
Интересно беше откритието на карбапенемазните гени *blaIMP-13-like* и *blaIMP-28/29-like* съответно с 99.87% нуклеотидна идентичност и 100% покритие. И трите гена бяха регистрирани от двете приложения ResFinder и AMRFinderPlus, респективно чрез суровите прочити и в асемблирания метагеном, но според AMRFinderPlus *blaIMP-28-like* не може категорично да се открие в метагенома (<50% покритие). Въпреки малките разминавания между приложенията успешно бяха асемблирани двата гена *blaIMP-13-like* и *blaIMP-29-like*. Те се детектираха в проба S7 (изходни рециклирани води) на два отделни контига с дължина съответно 1842 нб и 934 нб. На *Фигура №17* са демонстрирани как суровите прочити се налагаха върху референтните секвенции на IMP-13 и IMP-29. От *Фигура №17А* се вижда, че може да се касае за наличие на два алела *blaIMP* (13 и 13-like) по няколко вида прочити. На *Фигура №17В* асемблерът сглобява прочитите като един *blaIMP-29-like* ген, но се откриваха множество варианти на суровите прочитите в посочените позиции. Дълбочината на секвениране беше твърде ниска за по-обстоятелствен анализ. Не може да се отхвърлят варианти за наличие на повече от два алела на IMP.

От настоящите резултати две неща бяха отбелязани като особено любопитни, освен наличието на карбапенемазни гени: **а)** тяхната детекция само в пробите S10 – активна утайка и S7 – изходната рециклирана вода; и **б)** съпадението на подобни резултати с намерения *ген blaIMP-13-like* в р. Искър.

A. Contig with blaIMP-13-like



B.



Фигура №17. Консенсусни секвенции след налагане на суровите прочити към референтните секвенции на гените *blaIMP-13* и *blaIMP-29*. **A.** IMP-13 **B.** IMP-29

По отношение на **a)** се знае от литературния обзор, че ПСОВ са горещи точки за развитие и разпространение на AMP, но от *Фигура №16* се виждаше, че въпреки нормализирането на броя прочити спрямо дълбочината на секвениране, то по-голямата част от ГАР има значително по-малък бр. прочити в пробите активна утайка (S10) и рециклираната вода (S7), което най-често се интерпретира като намаляване на тяхното относително разпределение. Разбира се, това не е индикация за обратното твърдение, а и може да се дължи на стохастичния характер и композиционната структура на данните от метагеномиката. Но това относително намаляване помогна посредством heatmap графиката да се установи, че има ГАР, за които се наблюдава обратни тенденции. Например, при двата гена *blaIMP-13* и *blaIMP-28* беше наблюдавано точно обратното. Те не бяха открити в S8 (входящи води), а само в S10 (активна утайка, с ниско покритие), и с по-високо покритие в S7 (рециклирани води). Това означаваше, че продуцентът (или продуцентите) на двата гена *blaIMP-13-like* и *blaIMP-28/29-like* се намира и разпространява в ПСОВ гр. Стара Загора. Най-вероятно там микроорганизмът е изградил биофилм по повърхността на резервоарите и се явява първоизточникът на клетки, респективно ДНК и ГАР. Знае се, че биофилмите са устойчиви форми на съжителство на микробни сообщества, предоставят условия за

ХГП, което поставя въпросът дали и как гена (или гените) *blaIMP* се пренасят чрез ХГП на други микроорганизми. ГАР следващи същата тенденция бяха: *blaOXA-101*, *blaOXA-129*, *blaOXA-20*, *blaOXA-205*, *blaOXA-4*, *blaOXA-5*. За разлика от гена *blaIMP*, обаче, всички тези OXA варианти бяха открити още от входящите канални води (S8), но с по-малко прочити.

По отношение на **б)** след съпоставянето на двата контига от двата асемблирани метагенома заедно с референтната *blaIMP-13* беше установено, че те са различни. Генът от р. Искър не беше същият като открития в ПСОВ гр. Стара Загора. Въпреки това, учудващо е, че два гени, кодиращи карбапенемази се откриват в близки по характеристика проби. Логичен въпрос е дали и ПСОВ гр. Самоков също е проучен на ген *blaIMP*.

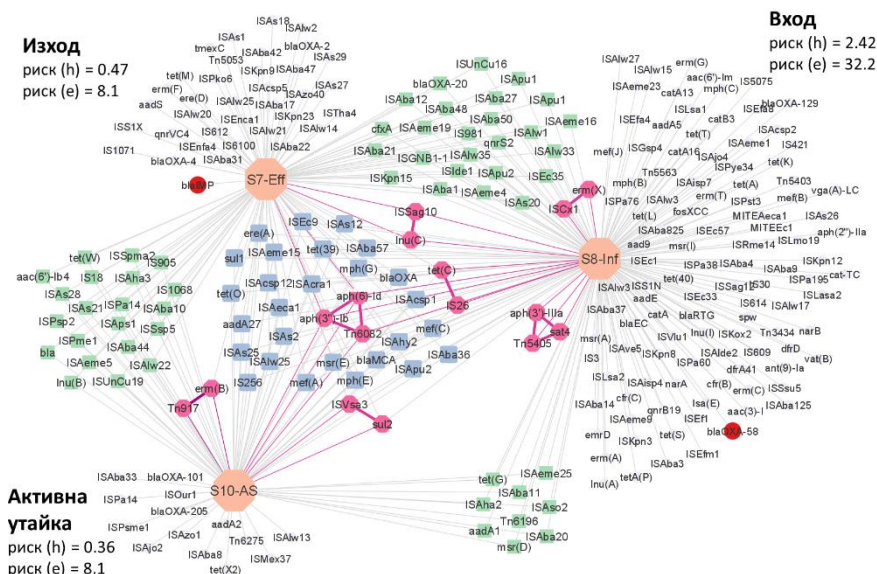
Беше извършен опит за култивиране на оригиналната проба от рециклираната вода върху селективни среди чрез карбапенемазна среда. Бяха селектирани *Acinetobacter*, *Pseudomonas*, *Stenotrophomonas* и др., но измежду тях не бяха открити продуценти на *blaIMP*. За това може да са допринесли няколко фактора. Не бяха направени опити за селекция посредством трето-генерационни цефалоспорини, понеже има IMP карбапенемази като IMP-4, придаващи ниски нива на МПК към карбапенеми [31], също така в зависимост от локализацията (плазмид или хромозома) и бр. копия на гена. В допълнение, оригиналната проба беше стара, въпреки че беше съхранявана на -80°C.

Не е изненадващо детектирането на уникални ГАР в активната утайка спрямо входящата вода, но е притеснително наличието на ГАР като карбапенемези от тип IMP или други, имащи тенденция към увеличаване на относителното количество. До голяма степен профилите на активната утайка и рециклирана вода може да си приличат една с друга и едновременно да са различни от входящата проба. Казано по друг начин, съставът на ГАР между трите основни точки на пробовземане може да варира. Това се дължи на уникалните качества и характеристики на утайката като отделна популация от микроорганизми, чията основна роля е да пречистват разтворените химикали, хранителни отпадъци, човешки отпадъци, неорганични твърди вещества. Микроорганизмите в утайката кодират собствен набор от ГАР, които съставляват по-голямата част краен състав, който се детектира в тези проби. Част от утайката от утайтелните резервоари (след като водата е била третирана с активна утайка и разделена) се връща обратно (още наричана върната утайка) и смесва към активната утайка (първичната смес между входяща вода и първична „прясна“ фракция на утайка). Връщането на утайка към активната утайка допълнително я съгъстява и концентрира твърдите вещества. Съотношението на върната утайка към активна утайка варира в различните ПСОВ и може да е от 25% до 75% спрямо първоначалната „прясна“ утайка. Тъй като този процес е непрестанен, активната утайка винаги представлява смес от „прясна“ (нова утайка), върната утайка и мръсна канална вода. По този начин, състава на ГАР в активната утайка може да варира

значително от състава на входящата мръсна вода поради факта, че там попадат концентрирани микроорганизми и ГАР от върната утайка. Така до голяма степен се запазва и проявява уникалния таксономичен и резистомен профил на утайката. На *Фигура №16* тези различия не са толкова силно изразени предвид фокуса върху клинично значимите ГАР. По-ниските нива на суровите прочити от клинично значимите ГАР пък се обясняват от една страна с намаляването на относителното количество на микроорганизми от входящата вода и голямата фракция на микроорганизми от утайката, които издърпват ресурс за секвениране.

4.4.3.5. Съвместна поява на ГАР и ПГЕ в предполагаеми плазмиди

Откриването и описването на ко-локализацията на ГАР и ПГЕ и по-специално върху плазмиди в ПСОВ осигурява изключително важни познания за механизмите на разпространение на резистентността. Тяхната ко-локализация подчертава потенциала за ХГП и може да се ползва като характерна черта за съпоставяне на резултатите и при оценка на способности за намаляване или ограничаване на AMP. В ПСОВ гр. Стара Загора най-голям набор ГАР и ПГЕ върху плазмиди бяха детектирани във входящата канална вода (S8) и активната утайка (S10) както е показано на *Фигура №18*.



Фигура №18. Съвместна поява на ГАР и ПГЕ върху потенциални плазмиди от ПСОВ гр. Стара Загора представени чрез мрежови анализ. ГАР и ПГЕ, срещани се само в две проби са отбелязани в зелено, в три проби - светло синьо. В лилав цвят са двойка ко-локализиращи елементи. Те са свързани с лилави ребра към дадена проба, което е индикация къде са открита. Гените *blaIMP* и *blaOXA-58* са оцветени в червено. Показани са само ГАР със 100% идентичност и ПГЕ с над 90% идентичност с изключение на *blaIMP*.

В S8, Tn5405 беше идентифициран заедно с гените *aph(3')-IIIa* (фосфотрансфераза за амикацин) и *sat4* (N-ацетилтрансфераза за стрептотрицин). Тн за първи път е бил регистриран в *S. aureus*. Tn5405 е 12 кб стафилококов композитен Тн, който впоследствие е открит в огромна част от изолатите на *S. aureus*, както и други стафилококи. В процес на рециклирането на водите, този Тн намаляваше в относително количество и не беше детектиран в следващите проби. Аналогично, *ermX* и *ISCx1* (ИЕ) бяха ко-локализирани само в S8. Тези два елемента се срещат в композитен Тн в род *Corynebacterium* и по-конкретно в *Corynebacterium xerosis*.

В S10 бяха регистрирани ISVsa3 (IS91-подобна транспозаза) заедно с *sul2* върху плазмиден контиг от малко над 4 кб. По литературни данни, произходът на ISVsa3 е от *Vibrio salmonicida*, като други чести гостоприемници са *E. coli*, *V. cholera* и *Salmonella enterica* subsp. *enterica* посредством плазмид. Например, *incC* плазмид, носещ *blaCMY-2* е идентифициран да носи в следната последователност гените *floR*, *tetA*, *aph(6)-Id*, *aph(3'')-Ib*, and *sul2* в изолати *S. enterica* от говеда. И двата елемента бяха открити и в трите проби, но само в активната утайка те бяха идентифицирани заедно на контиг. Допълнителни опити за подобряване на дължината на асемблираните контиги са необходими, за да се установи дали тези два елемента са постоянно асоциирани или тази находка е случайност. Tn917 и *ermB* бяха детектирани заедно само в S10. На последно място, Tn6082, носещ ГАР *aph(3'')-Ib* и *aph(6)-Id* за резистентност срещу стрептомицин бяха идентифицирани едновременно и в трите проби.

Стойността на риска за човешко здраве беше най-висока в пробата входяща канална вода S8. Изненадващо е, че рециклираната вода S7, която се зауства в р. Бидечка след ПСОВ гр. Стара Загора, показва по-висока стойност на риск от активната утайка S10. Към момента разработчиците на приложението MetaCompare2.0 са го настроили да оценява от 0 до 100, като стойности до 10 следва да се интерпретират като незамърсени проби или нискорискови. Смущаващо е, че всички настоящи проби в дисертационния труд попадат в този диапазон. Формулата оценява съотношението на нивата на клинично значими ГАР, патогенни контиги и ПГЕ по отделно както и в комбинации (ГАР-ПГЕ, ГАР върху контиг, ГАР върху контиг с патогенен произход, ГАР и ПГЕ върху контиг, и ГАР и ПГЕ върху контиг с патогенен произход) като на всеки от възможните комбинации присвоява различна степен на риск. Накрая се калкулира съотношението (относително количество) между тях и пълния брой от сурови прочити на метагенома.

В светлината на предстоящата директива на Европейската комисия за мониториране на АМР в големите ПСОВ [5], би трябвало да има начин за надеждно оценяване на замърсяването на отпадните води и поставянето им в дефинирани категории (напр. незамърсени, леко замърсени, силно замърсени). Формулата не взема под внимание вида на пробата или дали визуално е чиста или мръсна. Предполага се, че различните типове проби трябва да имат различни категории, в

които да се оценяват и сравняват. Към момента, линейната скала на приложението MC2 практически не разграничава активната утайка от рециклирана вода и канална вода. Формулата за калкулиране на този риск тепърва ще търпи повече усъвършенстване, защото много стъпки на стандартизиране са необходими да се приложат, за да се намалят отклоненията в крайната оценка.

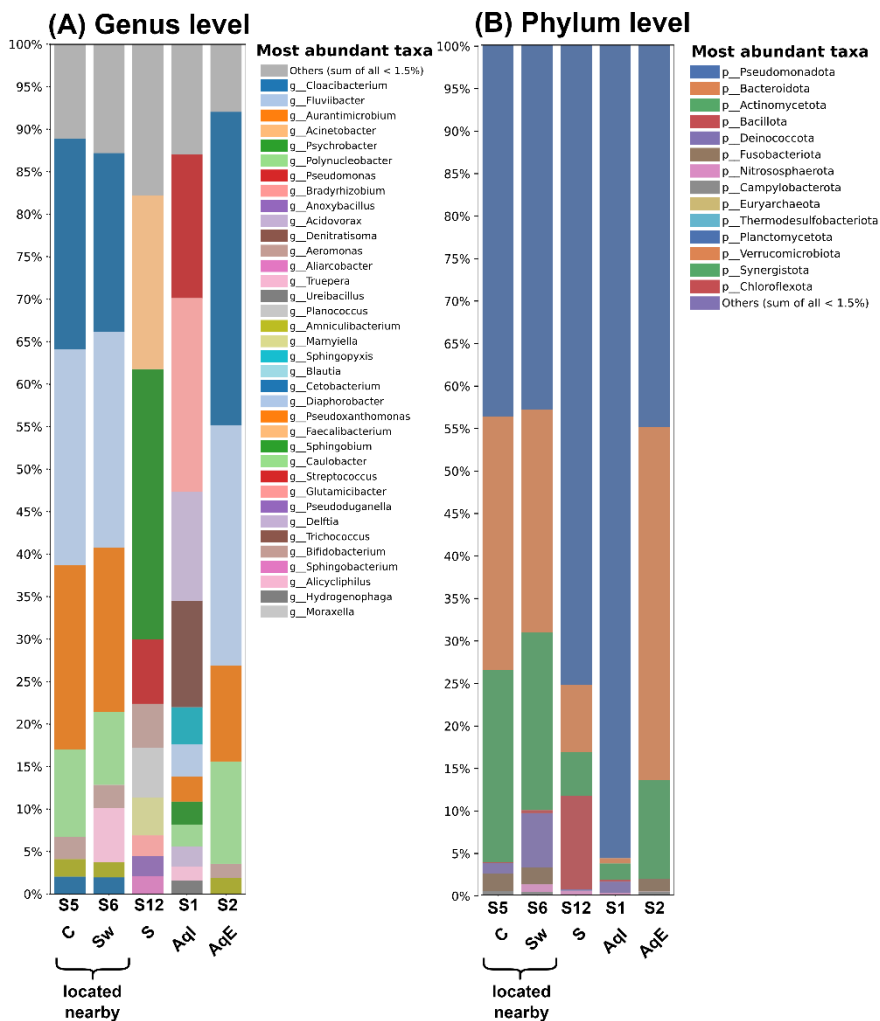
4.4.4. Анализи, базирани на сондажни (подпочвени) води в региона на гр. Стара Загора

Подпочвените води предоставят до 99% от наличната сладководна вода и са критичен воден източник в световен мащаб. През последните две десетилетия АМР се разпространява почти прикрито в околната среда посредством описаните пътища и механизми в настоящия труд. Съвсем логично е в отговор на редица доказателства да се търси и оценява наличието и степента на замърсяване в останалите части от околната среда. Наред с многото замърсители, установени в сладководни басейни (вкл. подпочвени води), като химикали, пестициди, пластмасови микро отпадъци и лекарства, така и антибиотиците са често откривани поради широката им употреба в животновъдството и медицината. Вследствие от отлагането на редица антибиотици като най-честите сулфаметоксазол, ципрофлоксацин и тетрациклини, редица БРА също са били открити и докладвани, които представляват риск за хората.

4.4.4.1. Таксономичен състав

Съдейки по сходството в таксономичния състав (Виж *Фигура №19*) на проби S5 и S6, както и тяхното близко разположение (~400 м), може да се каже, че тези два източника или са един и същ басейн подземна вода или са свързани помежду си.

С малки вариации в състава, всички четири сондажни проби имаха сходен таксономичен състав на ниво тип. Доминиращ беше тип *Pseudomonodota* (*Proteobacteria*) от близо 45% до 95% (S1-сондаж в аквапоника), следван от *Bacteroidota* и *Actinomycetota*. С изключение на *Actinomycetota* първите два типа обединяват най-често срещаните таксони в подпочвени води. На ниво род, интересна находка беше *Cloacibacterium*, който по литературни данни е регистриран в отпадъчни води, вливащи се в ПСОВ. Не се открива в човешки изпражнения и в рециклираните води след ПСОВ, което съответства на настоящите резултати. Двата идентифицирани вида в настоящите проби *C. caeni* (изолиран от активна утайка от ПСОВ) и *C. normanense*, оформят малко над 26% и 21% от таксономичния състав на съответно S5 и S6.

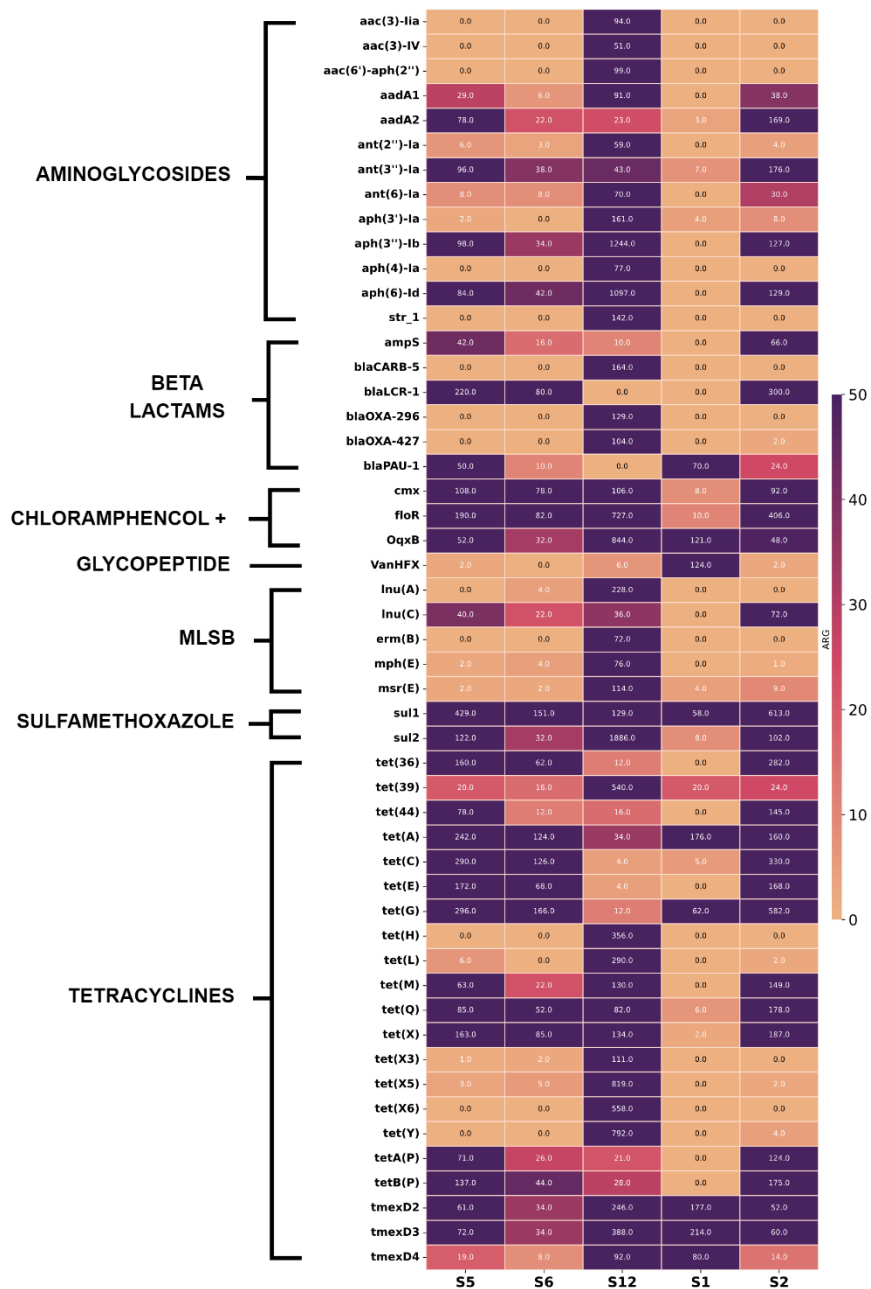


Фигура №19. Барплот графики с таксономичен състав на ниво род (A) и тип (B), изобразяващи относителното изобилие на най-разпространените таксони във всяка от пробите. S5-C - сондаж при кравеферма, S6-Sw – сондаж при свинеферма. S5 и S6 са разположени на физическо 200м разстояние. S12-S – овцеферма, S1-AqI и S2-AqE са съответно сондаж при аквапоника и утаителна шахта след аквапониката. Нискоразпространените таксони (< 0.1 %) и таксоните, открити в по-малко от 2 проби, бяха филтрирани. Таксони с изобилие под 1.5% бяха групирани в група "Others".

4.4.4.2. Детектиране на ГАР

Клинично значимите ГАР, които бяха открити в пробите от сондажна вода са визуализирани на *Фигура №20*. Измежду тях най-голямо разнообразие беше регистрирано в пробата S12, която беше от село Християново. Това беше единствената локация, при която пробовземането беше извършено през маркуч, който отвежда водата от сондажа до най-близката сграда. На външен вид тя имаше жълтеникав и леко мътен цвят. Поради това, само около 350 мл бяха филтрирани през 0.2 μ m филтър. Въпреки това, в този набор проби не бяха регистрирани карбапенемази или БЛРС.

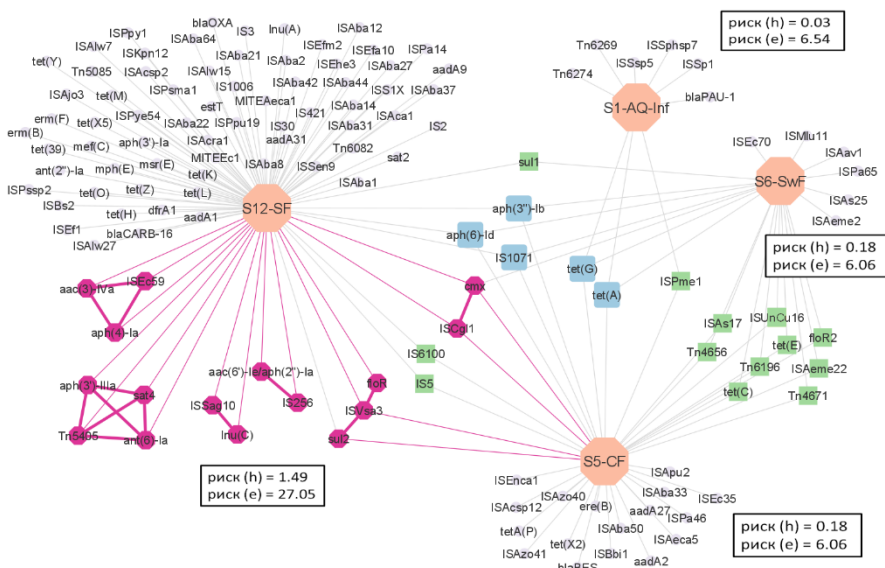
Най-голямо разнообразие беше наблюдавано при ГАР към тетрациклини. Не е случайно тяхното широко присъствие в околната среда като океани и морета, географски много отдалечени локации спрямо населяваните от човека. Тетрациклиновите антибиотици са група антибиотици, използвани за ветеринарни цели, за терапия на хора и за селскостопански цели в продължение на десетилетия, за разлика от други по-нови молекули и класове. Поради широката им употреба тетрациклиновите антибиотици са се оказали повсеместно разпространени съединения, открити в различни екологични ниши. При прием на тетрациклини повече от 70% от приемната доза се отделя и освобождава в активна непроменена форма в околната среда. В допълнение на това хидрофилния характер и ниското ниво на променливост на техните молекули ги правят значителна по-устойчиви във водната среда. Всички тези фактори заедно допринасят за дългогодишната селекция на микроорганизми резистентни към тетрациклини, които могат да предизвикат неблагоприятни последици за човешкото здраве, като увеличават риска от някои инфекции.



Фигура №20. Heatmap графика от нормализираните прочити, които успешно се налагаха върху ГАР с клинично значение. Всички бяха открити с минимум 60% покритие и 60% идентичност. Гените са подредени въз основа на класа антимикробни средства, към които придават резистентност. Всички гени с $n=50+$ прочита са оцветени в един и същи цвят с цел визуализация. MLSB се отнася до ГАР, които придават резистентност към един или няколко от следните антибиотици: еритромицин, азитромицин, телитромицин, линкозамид, стрептограмин В/А, линкомицин, клиндамицин, хинупристин и далфопристин. Групата Chloramphenicol + включва резистентност към: амфеникол, хлорамфеникол, оксазолидинон, линкозамид, стрептограмин А, флорфеникол, клиндамицин, линезолид, ципрофлоксацин, триметоприм.

4.4.4.3. Съвместна поява на ГАР, ПГЕ върху плазмиди

На *Фигура №21* е представено носителството на ГАР и ПГЕ върху плазмиди. Най-голям спектър от ГАР и ПГЕ, както и случаи на ко-локализация бяха регистрирани в подпочвените води S12 в с. Християново. Респективно, в тази проба рисковият фактор беше най-висок. Автори, които също изследват подпочвени води и регистрират ГАР, сред които най-много тетрациклинови варианти, заключват че калкулираният риск в настоящите си граници не представляват директна заплаха за човешкото здраве [32]. Само сулфаметоксазол, ципрофлоксацин и еритромицин имат умерени до ниски нива на риск за водните обитатели. Единствено консумацията на подпочвени води, замърсени с ГАР и БРА би представлявала риск за развитието на AMP при хората [32].

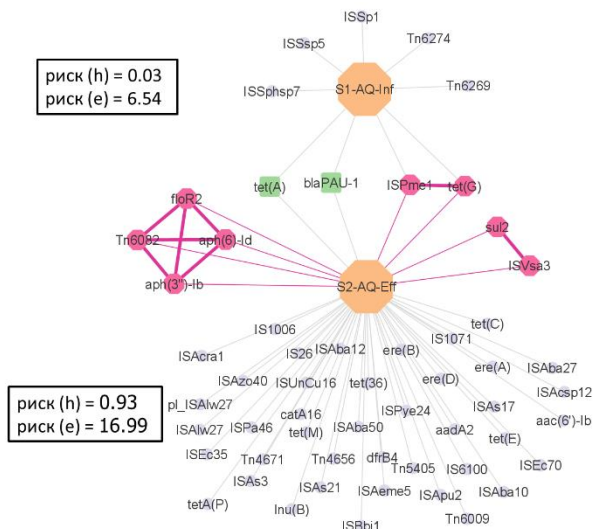


Фигура №21. Мрежови анализ на ГАР открити върху потенциални плазмиди от сондажни води от региона на гр. Стара Загора. ГАР и ПГЕ, срещащи се само в две проби, са отбелязани в зелено. В светло синьо са отбелязани ГАР и ПГЕ, които се срещат и в трите проби. В розов цвят и свързани с розови ребра са ПГЕ и ГАР, които са локализиращи на един контиг. Показани са само ГАР със 100% идентичност и ПГЕ с над 90% идентичност.

На *Фигура №22* е представена динамиката на ГАР, ПГЕ и плазмиди, които се регистрират във входящата сондажна вода (S1) и в изходяща вода от утайтелна шахта (S2) в аквапоника за едновременно отглеждане на риби и зелена салата. Тн Tn6082, носещ *aph(6)-Id* и *aph(3'')-Ib*, беше регистриран в утайтелната шахта. Той също така беше регистриран и в пробите от ПСОВ гр. Стара Загора, както и в река Искър, което говори за неговото широко разпространение. В настоящата проба (S2), вероятно секвенционното покритие на този участък е по-високо, което е от своя страна е генерирало по-дълъг контиг по време на асемблирането. В резултат, Tn6082 носи и допълнителен ГАР *flor2*. Също така, ISVsa3 и *sul2* бяха локализиращи заедно в проба S2, аналогично на проба S10 (активна утайка от ПСОВ гр. Стара Загора).

Необходимо е да се направи корелационен анализ между ГАР, ПГЕ и откритите таксономични единици при всички проби, за да се получи информация кои

микроорганизми биха били потенциален гостоприемник на тези ко-локализиращи елементи. Това ще бъде обект на допълнителна работа извън настоящия дисертационен труд.



Фигура №22. Мрежови анализ върху динамиката на ГАР и ПГЕ от сондажна вода (S1-AQ-Inf), която захранва системата на аквапониката и преминала през цялото трасе до утайтелна шахта (S2-AQ-Eff). В зелено са отбелязани ГАР и ПГЕ, които се срещат и в двете проби. В розов цвят и свързани с розови ребра са ПГЕ и ГАР, които са локализиращи на един контиг. Показани са само ГАР със 100% идентичност и ПГЕ с над 90% идентичност.

5. ИЗВОДИ

1. Разработения от нас *in-house* протокол за изолиране на ДНК предоставя съпоставими резултати с установени търговски набори, водейки до ниски нива на отклонения в таксономичния състав на микробен стандарт, и е подходящ за метагеномни проучвания.
2. От шестте изпитани публично достъпни бази данни за идентификация по гена 16S рPHK, GSR и GTDB са най-надеждни за провеждане на 16S метагеномно проучване.
3. Методиката Флокуляция със обезмаслено мляко не е подходяща за концентриране на бактериална биомаса за целите на метагеномни проучвания, поради способността на някои бактерии в хода на инкубацията да се размножават, метаболизирайки обезмасленото мляко.
4. Заустването на рециклираната вода от ПСОВ гр. Самоков води до завишени относителни нива на гени резистентност към клинично значими антимикробни средства (макролиди, тетрациклини, бета-лактами, карбапенеми, сулфонамиди и др.) в река Искър.
5. Проучените ПСОВ не са достатъчно ефективни за пречистване на резистентни към антибиотици бактерии.
6. Вследствие на проучената динамика на гените за резистентност беше доказано, че ПСОВ гр. Стара Загора е първоизточник на ген за карбапенемаза *blaIMP*.
7. Значителна част от установените гени за резистентност се асоциираха с подвижни генетични елементи повишавайки техния потенциал за хоризонтален генетичен пренос.

6. ЗАКЛЮЧЕНИЕ

Антимикробната резистентност на клинично значими бактерии представлява широко разпространено глобално предизвикателство в здравеопазването, което през последните две десетилетия е назряващ проблем и в околната среда. Вследствие от приема на антибиотици от хора и животни, или от промишлена производствена дейност, част от тях се отделя в околната среда или в градската канализация. От там, те достигат до пречиствателните станции за отпадъчни, които играят ролята на фуния, в чиито тесен край всички микроорганизми и антибиотици се концентрират и взаимодействат помежду си, превръщайки станциите в горещи точки за възникване и разпространение на антимикробна резистентност. Този процес се опосредства от неспособността на пречиствателните станции да пречистват напълно от микроорганизми и остатъчни антибиотици.

През юли 2022 г. Европейската комисия заедно с държавите членки, определи антимикробната резистентност като една от трите най-големи заплахи за здравето от основен приоритет. Това доведе и до множество политически решения като разработване на стратегии за борба срещу антимикробната резистентност, както и насърчаване на подхода „Едно Здраве“. Сред тях са предложението за преразглеждане на действащите списъци със замърсителите на повърхностни и подземни води, както и да се имплементира национално мониториране за антимикробната резистентност в градските пречиствателни на станции за отпадъчни води, обслужващи над 100,000 души [5]. Като Национална Референтна Лаборатория по „Контрол и Мониториране на Антимикробната Резистентност“, наша роля е да изградим капацитет за провеждане на такива метагеномните проучвания.

Целта на настоящия труд „Да се проучат разпространението и вида на детерминантите на антимикробна резистентност, подвижните генетични елементи и бактериалното биоразнообразие в проби от околна среда и потенциалния риск за здравето на човека“ е съпроводена с няколко отделни задачи. Изпълнявайки всички заложи задачи, смятаме, че целта са постигната. Считаме, че беше отбелязан голям напредък в областта на клиничната микробиологията и микробиологията на околната среда за територията на България.

Това е първото метагеномно проучване, което дава оценка върху ефекта на пречиствателната станция за отпадъчни води на гр. Самоков върху резистома, мобилома и микробиома на река Искър, основен сладководен източник на питейна вода за Софийска област. Идентифицирани бяха както фекални замърсители, така и придобити гени за антимикробна резистентност, включително такива с клинично значение като карбапенемазните гени *blaOXA-58* и *blaIMP*. Това ясно подчертава въздействието на антропогенното замърсяване върху сладководните екосистеми посредством пречиствателните станции за отпадъчни води. Заедно с това, бяха идентифицирани потенциални гостоприемници на редица гени за резистентност,

както и подвижни генетични елементи, подкрепящи хипотезата за неконтролирано разпространение на антимикробна резистентност в околната среда.

В пречиствателна станция за отпадъчни води на гр. Стара Загора бяха идентифицирани още по-голямо разнообразие от гени за антимикробна резистентност. Освен резистентност към макролиди, тетрациклини, сулфонамиди и аминогликозиди, там бяха открити гени за широкоспектърни бета-лактамази и карбапенемази. Любопитно беше установяването на двете карбапенемази *blaOXA-58* и *blaIMP* в пробите от р. Искър и в пробите от пречиствателната станция на гр. Стара Загора. Значим резултат беше, че първоизточникът на *blaIMP* беше самата пречиствателна станция на гр. Стара Загора, понеже генът не беше открит във входящите канални води, а само в активната утайка и рециклираните води. На последно място бяха проучени и подпочвени води в региона на гр. Стара Загора, в които бяха подробно описани състава на микробиома, резистома и мобилома.

Заедно с това, може да отбележим напредъка с методичен характер. Беше разработен и оптимизиран *in-house* протокол за изолиране на ДНК за бактериална метагеномика, който показва конкурентни резултати в сравнение с комерсиалните комплекти, които са оптимизирани за минимално отклонение при метагеномни проби. Чрез сравнителен анализ бяха проучени особеностите, предимствата и недостатъците на шест различни 16S рРНК бази данни за таксономично идентифициране. На последно място беше проучено приложението на методиката Флокулация с обезмаслено мляко за концентриране на биомаса при бактериална метагеномика. Тази методика в настоящия си протокол изменяше значително таксономичния състав на пробите и е неприложима за тези цели.

Бъдещи планове включват разширяване на проучените пречиствателни станции за установяване на общите нива на антимикробна резистентност за страната, разширяване на локациите за пробовземане както в станциите, така и преди и след. Това би характеризирало детайлно динамиката на резистентността и къде точно са най-горещите точки, съответно да се предложат механизми за намаляване степента на възникване и разпространение на резистентността. Също така, беше планирано да се изследват остатъчни антибиотични молекули във пробите и да се установи съществуват ли асоциации с резистентността. Бъдещи проучвания биха могли да включват използването на културелни методи за установяването на носителите на *blaOXA-58* и по-важно *blaIMP*.

7. ПРИНОСИ

Приноси с фундаментален характер:

1. Беше установен и характеризиран ефектът на метода *флокуляция със обезмаслено мляко* като алтернатива на вакуум филтрирането за концентриране на бактериална биомаса за метагеномни проучвания

Приноси с научни-приложен характер:

2. Значително е обогатена информацията за разпространението на AMP в подпочвени води и река Искър в България чрез характеризиране на резистома и мобилома
3. Значително е обогатена информацията за разпространението на AMP в ПСОВ и влиянието върху резистома и мобилома на рециклираните води върху приемащия обект
4. Установено е наличие на карбапенемазни гени в и след ПСОВ, като за едната от ПСОВ беше установено, че тя е първоизточника на гена *blaIMP*
5. Чрез сравнителни анализи бяха характеризирани отклоненията, които те внасят върху таксономичния състав на метагеномни проби, от шест бази данни за 16S идентификация и от пет протокола за изолиране на ДНК
6. Това е първото всеобхватно метагеномно проучване, насочено към документиране на бактериалното таксономично разнообразие на сладководна река в България, и ефектите ПСОВ гр. Самоков върху резистома и мобилома

Приноси с методичен характер:

1. Оптимизиран и имплементиран е протокол за генериране секвенционни библиотеки за 16S rRNA метагеномно проучване.
2. Разработен, оптимизиран и имплементиран е *in-house* протокол за изолиране на ДНК за бактериална метагеномика.

8. СПИСЪК НА ЦИТИРАНА ЛИТЕРАТУРА

- [1] Murray CJ, Ikuta KS, Sharara F, Swetschinski L, Robles Aguilar G, Gray A, et al. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *Lancet Lond Engl* 2022;399:629–55. [https://doi.org/10.1016/S0140-6736\(21\)02724-0](https://doi.org/10.1016/S0140-6736(21)02724-0).
- [2] Antimicrobial resistance in the EU/EEA (EARS-Net) - Annual epidemiological report for 2021 2022. <https://www.ecdc.europa.eu/en/publications-data/surveillance-antimicrobial-resistance-europe-2021> (accessed September 21, 2024).
- [3] Singer AC, Shaw H, Rhodes V, Hart A. Review of Antimicrobial Resistance in the Environment and Its Relevance to Environmental Regulators. *Front Microbiol* 2016;7:1728. <https://doi.org/10.3389/fmicb.2016.01728>.
- [4] Sun C, Li W, Chen Z, Qin W, Wen X. Responses of antibiotics, antibiotic resistance genes, and mobile genetic elements in sewage sludge to thermal hydrolysis pre-treatment and various anaerobic digestion conditions. *Environ Int* 2019;133:105156. <https://doi.org/10.1016/j.envint.2019.105156>.
- [5] Proposal for a revised Urban Wastewater Treatment Directive - European Commission n.d. https://environment.ec.europa.eu/publications/proposal-revised-urban-wastewater-treatment-directive_en (accessed January 23, 2024).
- [6] de Bruin OM, Chiefari A, Wroblewski D, Egan C, Kelly-Cirino CD. A novel chemical lysis method for maximum release of DNA from difficult-to-lyse bacteria. *Microb Pathog* 2019;126:292–7. <https://doi.org/10.1016/J.MICPATH.2018.11.008>.
- [7] Harkes P, Suleiman AKA, van den Elsen SJJ, de Haan JJ, Holterman M, Kuramae EE, et al. Conventional and organic soil management as divergent drivers of resident and active fractions of major soil food web constituents. *Sci Rep* 2019 91 2019;9:1–15. <https://doi.org/10.1038/s41598-019-49854-y>.
- [8] Abraham D, Mohan VR, Kang G. Skimmed Milk Flocculation Technique for Waste Water 2021. <https://doi.org/10.17504/PROTOCOLS.IO.BUZZMNX46>.
- [9] Takahashi S, Tomita J, Nishioka K, Hisada T, Nishijima M. Development of a Prokaryotic Universal Primer for Simultaneous Analysis of Bacteria and Archaea Using Next-Generation Sequencing. *PLOS ONE* 2014;9:e105592. <https://doi.org/10.1371/JOURNAL.PONE.0105592>.
- [10] Zymo-Research/miqScore16SPublic 2023.
- [11] Zymo-Research/miqScoreShotgunPublic 2024.
- [12] Donchev D. maddne/MIQ-calc-from-OTU-tables 2024.
- [13] Lu J, Rincon N, Wood DE, Breitwieser FP, Pockrandt C, Langmead B, et al. Metagenome analysis using the Kraken software suite. *Nat Protoc* 2022 1712 2022;17:2815–39. <https://doi.org/10.1038/s41596-022-00738-y>.
- [14] Lin H, Peddada SD. Analysis of compositions of microbiomes with bias correction. *Nat Commun* 2020;11:3514. <https://doi.org/10.1038/s41467-020-17041-7>.
- [15] Rumi MA, Oh M, Davis BC, Juvekar A, Brown CL, Vikesland PJ, et al. MetaCompare 2.0: Differential ranking of ecological and human health resistome risks. *bioRxiv* 2024:2024.01.17.576132. <https://doi.org/10.1101/2024.01.17.576132>.
- [16] Calgua B, Mengewein A, Grunert A, Bofill-Mas S, Clemente-Casares P, Hundesa A, et al. Development and application of a one-step low cost procedure to concentrate viruses from seawater samples. *J Virol Methods* 2008;153:79–83. <https://doi.org/10.1016/J.JVIROMET.2008.08.003>.
- [17] Philo SE, Ong AQW, Keim EK, Swanson R, Kossik AL, Zhou NA, et al. Development and Validation of the Skimmed Milk Pellet Extraction Protocol for SARS-CoV-2 Wastewater Surveillance. *Food Environ Virol* 2022;14. <https://doi.org/10.1007/S12560-022-09512-5>.
- [18] Gonzales-Gustavson E, Cárdenas-Youngs Y, Calvo M, da Silva MFM, Hundesa A, Amorós I, et al. Characterization of the efficiency and uncertainty of skimmed milk flocculation for the simultaneous concentration and quantification of water-borne viruses, bacteria and protozoa. *J Microbiol Methods* 2017;134:46–53. <https://doi.org/10.1016/J.MIMET.2017.01.006>.
- [19] Hiergeist A, Ruelle J, Emler S, Gessner A. Reliability of species detection in 16S microbiome analysis: Comparison of five widely used pipelines and recommendations for a more

- standardized approach. PLOS ONE 2023;18:e0280870. <https://doi.org/10.1371/JOURNAL.PONE.0280870>.
- [20] Knudsen BE, Bergmark L, Munk P, Lukjancenko O, Priemé A, Aarestrup FM, et al. Impact of Sample Type and DNA Isolation Procedure on Genomic Inference of Microbiome Composition. mSystems 2016;1:e00095-16. <https://doi.org/10.1128/mSystems.00095-16>.
- [21] Yang L, Chen J. A comprehensive evaluation of microbial differential abundance analysis methods: current status and potential solutions. Microbiome 2022;10:130. <https://doi.org/10.1186/s40168-022-01320-0>.
- [22] Yuan L, Dai H, He G, Yang Z, Jiao X. Invited review: Current perspectives for analyzing the dairy biofilms by integrated multiomics. J Dairy Sci 2023;106:8181–92. <https://doi.org/10.3168/JDS.2023-23306>.
- [23] Cacace D, Fatta-Kassinos D, Manaia CM, Cytryn E, Kreuzinger N, Rizzo L, et al. Antibiotic resistance genes in treated wastewater and in the receiving water bodies: A pan-European survey of urban settings. Water Res 2019;162:320–30. <https://doi.org/10.1016/J.WATRES.2019.06.039>.
- [24] Breton-Deval L, Sanchez-Flores A, Juárez K, Vera-Estrella R. Integrative study of microbial community dynamics and water quality along The Apatlaco River. Environ Pollut 2019;255:113158. <https://doi.org/10.1016/J.ENVPOL.2019.113158>.
- [25] Chopyk J, Nasko DJ, Allard S, Bui A, Pop M, Mongodin EF, et al. Seasonal dynamics in taxonomy and function within bacterial and viral metagenomic assemblages recovered from a freshwater agricultural pond. Environ Microbiomes 2020;15:1–16. <https://doi.org/10.1186/S40793-020-00365-8/FIGURES/4>.
- [26] Ботева НИ, Ботева НИ. Разнообразие и биотехнологичен потенциал на термофилни микроорганизми от български горещи извори: дисертационен труд за присъждане на образователна и научна степен “доктор” област на висше образование 4. Природни науки, математика и информатика, професионално направление: 4.3 “Биологически науки”, научна специалност 01.06.12 “Микробиология.” [Н. Ботева], 2021.
- [27] Iliev I, Yahubyan G, Marhova M, Apostolova E, Gozmanova M, Gecheva G, et al. Metagenomic profiling of the microbial freshwater communities in two Bulgarian reservoirs. J Basic Microbiol 2017;57:669–79. <https://doi.org/10.1002/JOBM.201700137>.
- [28] Tsvetanova Z, Najdenski H. Antimicrobial Resistance of Heterotrophic Bacteria and Enterobacteriaceae Inhabiting an Anthropogenic-Affected River Stretch in Bulgaria. Process 2023 Vol 11 Page 2792 2023;11:2792. <https://doi.org/10.3390/PR11092792>.
- [29] Panaiotov S, Simeonovski I, Levterova V, Karamfilov V, Brankova N, Tankova K, et al. Two-Year Monitoring of Water Samples from Dam of Iskar and the Black Sea, Bulgaria, by Molecular Analysis: Focus on Mycobacterium spp. Int J Environ Res Public Health 2015 Vol 12 Pages 7430-7443 2015;12:7430–43. <https://doi.org/10.3390/IJERPH120707430>.
- [30] Conco T, Kumari S, Awolusi OO, Allam M, Ismail A, Stenström TA, et al. Profiling of emerging pathogens, antibiotic resistance genes and mobile genetic elements in different biological wastewater treatment plants. J Environ Chem Eng 2022;10:107596. <https://doi.org/10.1016/j.jece.2022.107596>.
- [31] Liu Z, Li J, Wang H, Xia F, Xia Y, Wang H, et al. Clonal transmission of blaIMP-4-carrying ST196 Klebsiella pneumoniae isolates mediated by the IncN plasmid in China. J Glob Antimicrob Resist 2024;38:116–22. <https://doi.org/10.1016/j.jgar.2024.05.002>.
- [32] Zainab SM, Junaid M, Xu N, Malik RN. Antibiotics and antibiotic resistant genes (ARGs) in groundwater: A global review on dissemination, sources, interactions, environmental and human health risks. Water Res 2020;187:116455. <https://doi.org/10.1016/j.watres.2020.116455>.

9. СПИСЪК НА НАУЧНИ ТРУДОВЕ, СВЪРЗАНИ С ДИСЕРТАЦИОННИЯ ТРУД, И ПРОЕКТИ

Публикации:

Donchev, D.; Ivanov, I.N.; Stoikov, I.; Ivanova, M. Metagenomic Investigation of the Short-Term Temporal and Spatial Dynamics of the Bacterial Microbiome and the Resistome Downstream of a Wastewater Treatment Plant in the Iskar River in Bulgaria. *Microorganisms* 2024, 12, 1250. <https://doi.org/10.3390/microorganisms12061250>, Q2, IF 4.1

Donchev, D.; Stoikov, I.; Diukendjieva, A.; Ivanov, I.N. Assessment of Skimmed Milk Flocculation for Bacterial Enrichment from Water Samples, and Benchmarking of DNA Extraction and 16S rRNA Databases for Metagenomics. *Int. J. Mol. Sci.* 2024, 25, 10817. <https://doi.org/10.3390/ijms251910817>, Q1, IF 4.9

Участия на научни форуми:

"Метагеномно изследване на краткосрочната времева и пространствена динамика на бактериалния микробиом и резистома в река Искър след пречиствателна станция за отпадъчни води"

Доклад представен на *XXII Национален Конгрес по Клинична Микробиология и Инфекции на Българската Асоциация на Микробиолозите*, гр. София, България, проведен на 13 - 15 Септември 2024 г.

„Variations in bacterial taxonomic composition of river water treated with skimmed milk flocculation and vacuum filtration“

Постер представен на конгрес *FEMS2023*, гр. Хамбург, Германия. проведен на 9 - 13 Юли 2023 г.

Участия в други проекти:

„Неинвазивни маркери за възпаление при пациенти с ринит с или без придружаваща бронхиална астма“, ръководител д-р Диана Христова

“Имунен дисбаланс и дефицити“ към Програмата за стратегически изследвания и иновации за развитие на МУ – Пловдив (ПСНИИР-МУП)“

“Молекулярно-вирусологичен анализ на въведения и разпространен нововъзникнал пандемичен вирус SARS-CoV-2 в България, чрез новогенерационно секвениране и комбиниран епидемиологичен и филогенетичен анализ“, ръководител доц. Ивайло Алексиев, дб

Изследовател по РП1 и РП2 по проект BG05M2OP001-1.002-0001-C04

„Фундаментални, транслиращи и клинични изследвания в областта на инфекциите и инфекциозната имунология“, НЦЗПБ, София

10. БЛАГОДАРНОСТИ

Изказвам своята благодарност на научния си ръководител доц. Иван Иванов, както и на ръководството на НЦЗПБ, за оказаната ми възможност да работя върху настоящия дисертационен труд. Благодаря за помощта, полезните съвети, напътствия и проявеното разбиране.

Благодаря на целия екип на Националната референтна лаборатория „Контрол и Мониториране на Антибиотичната резистентност“ за оказаната подкрепа и за мъдрите житейски съвети.

Това проучване беше подкрепено от Националния фонд за научни изследвания на Република България чрез проект КП-06-Н23/5 и Европейския фонд за регионално развитие чрез Оперативна програма "Наука и образование за интелигентен растеж" 2014-2020; проект BG05M2OP001-1.002-0001-C04 "Фундаментални, транслиращи и клинични изследвания в областта на инфекциите и инфекциозната имунология"; както и от асоциация Биокампус София.

11. ABSTRACT

Introduction: Antimicrobial resistance (AMR) represents a widespread global challenge in both clinical settings and the environment. As a result of the imprudent use of antibiotics in the human and animal sectors in Bulgaria, AMR and resistant bacteria (RB) have been silently spreading in the environment via the sewage and the Waste Water Treatment Plants (WWTP). AMR and RB pose a certain public health risk in environmental sources such as water, soil, and air through water sports, open swimming in natural water bodies, drinking from natural freshwater sources without treatment, as well as agriculture, irrigation with untreated water in crop production, and inhalation of aerosols. The EU commission has recently proposed the implementation of national monitoring of hotspots or collection points for AMR development, such as WWTPs serving more than 100,000 people. As part of the National Reference Laboratory for Control and Monitoring of Antimicrobial Resistance, NCIPD in Bulgaria, our role is to build the capacity to conduct such metagenomic studies.

Aim: Investigation of antimicrobial resistance genes (ARG), mobile genetic elements (MGE), and their prevalence and bacterial biodiversity in environmental samples and assessment of the potential risk to human health. Furthermore, by benchmarking, we studied the applicability of skimmed milk flocculation (SMF) for bacterial enrichment in metagenomics, an adapted in-house DNA extraction protocol, and six 16S rRNA databases (16S-DB).

Materials and Methods: Surface water samples from two rivers were treated with SMF and vacuum filtration (VF) and subjected to amplicon or shotgun metagenomics. A microbial community standard was subjected to five DNA extraction protocols, taxonomical identification with six different 16S-DBs, and evaluation using the Measurement Integrity Quotient (MIQ) score. For AMR studies, eight surface water samples were collected from River Iskar right after WWTP Samokov, three samples (influent, active sludge, and effluent) from WWTP Stara Zagora, two samples from a small local WWTP within a poultry house, and four underground water samples, located within swine, sheep, cow farms, and an aquaponics. For aquaponics, an additional water sample was collected at the end of the irrigation path. Samples were subjected to either 16S V3-V4 amplicon or shotgun metagenomic sequencing on MiSeq V3 (2 x 300 bp) or NextSeq Mid/High Output (2 x 150 bp).

Results and Discussion: In River Iskar, a high relative abundance of the genes *erm(B)*, *erm(F)*, *mph(E)*, *msr(E)* (macrolides); *tet(C)*, *tet(O)*, *tet(W)*, *tet(Q)* and *tet(X)* (tetracyclines); *sul1* and *sul2* (sulphonamides); and *cfxA3*, *cfxA6* (beta-lactams) were detected, with an observed trend of increased presence in the later sampling dates and locations closer to the WWTP. Notably, genes conferring resistance to carbapenems *blaOXA-58* and *blaIMP-33*-like were identified. Co-occurrence analysis of ARGs and MGE on putative plasmids revealed few instances, such as *blaOXA-58* and ISAlw25, Tn917 with its internal *ermB*, and Tn6082, carrying *aph(6)-Id* и *aph(3'')-Ib*. In total, 29 metagenome-assembled genomes were recovered, with only a few harbouring ARGs. This study enhances our understanding of how

a WWTP could influence the resistome and mobilome of the receiving freshwater river, underscoring the importance of continued ARG monitoring within the environment.

At the WWTP Stara Zagora a few ESBL genes were discovered, including *blaOXA-1*, *blaOXA-10*, *blaOXA-129*, *blaGES-5* in addition to two carbapenemases *blaOXA-58* and *blaIMP13*-like. While *blaOXA-58* was identified in the influent water, *blaIMP13*-like were detected only in the active sludge and in the effluent recycled water, indicating that WWTP Stara Zagora is a primary source of the carbapenemase gene *blaIMP*. Numerous ARGs conferring resistance to tetracyclines, macrolides and other beta-lactams were common for the WWTP. Examples of ARG-MGE co-localization included ISVsa3 with *sul2*, Tn917 with *ermB*, Tn6082 with *aph(3'')-Ib* и *aph(6)-Id*, among others. Except for one sample, very few ARGs and MGEs were detected in the underground water samples. Co-localization of ARG-MGE on putative plasmids were also characterized. Across all shotgun metagenomic samples, the estimated human health risk score ranged from 0.18 (freshwater sample from River Iskar) to 2.42 (for influent sewage from WWTP Stara Zagora), which was considered low according to MetaCompare2.0. However, these results should be interpreted with caution.

According to our benchmarking studies, in SMF-treated samples, skimmed milk was metabolized by members of lactic acid bacteria or genera such as *Polaromonas*, *Macroccoccus*, and *Agitococcus*, resulting in a significant increase ($p < 0.5$) in relative abundance up to 5.0 log fold change compared to VF. This rendered SMF inapplicable for bacterial microbiome studies. The best-performing DNA extraction protocols were FastSpin Soil, the in-house method, and EurX. All 16S-DBs produced comparable MIQ scores within each DNA extraction kit, ranging from 61–66 for ZymoBIOMICs and up to 80–82 for FastSpin. DNA extraction kits exerted more bias toward taxonomic composition than the choice of 16S-DBs.

Conclusion: Further studies are needed to provide a broader snapshot of the current AMR spread within Bulgarian WWTPs and to enhance our understanding of freshwater microbial community dynamics and antibiotic resistance profiles.