

ОДЕСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ імені І. І. МЕЧНИКОВА
Біологічний факультет
Кафедра мікробіології, вірусології та біотехнології

**Дипломна робота
бакалавра**

на тему: «Дослідження кластерів генів вторинних метаболітів штаму
Streptomyces ambofaciens Mut7 ізольованого з Чорного моря»

*Study of gene clusters of secondary metabolites of Streptomyces ambofaciens Mut7
strain isolated from the Black Sea*

Виконала: студентка IV курсу
денної форми навчання
спеціальність 162 Біотехнології та
біоінженерія
Голубєва Ілона Юріївна

Науковий керівник:
кандидат біологічних наук, доцент
Васильєва Наталія Юріївна

Рецензент:
кандидат біологічних наук, доцент
Назарчук Юлія Сергіївна

Рекомендовано до захисту:
Протокол засідання кафедри
№ від « » 2023 р.

Захищено на засіданні ЕК №
Протокол № _____ від «__» _____ р.
Оцінка _____ / _____ / _____
(за національною шкалою, шкалою ECTS, бал)

Завідувач кафедри

(підпис) Філіпова Т.О.
(прізвище та ініціали)

Голова ЕК

(підпис)

Ямборко Г.В.
(прізвище та ініціали)

Одеса – 2023

АНОТАЦІЯ

В роботі був проведений біоінформатичний аналіз геному *Streptomyces ambofaciens* Myt7 ізольованого з Чорного моря для дослідження кластерів генів вторинних метаболітів та з'ясування перспективи використання цих вторинних метаболітів. Дослідження наявності кластерів генів було проведено з використанням програми antiSMASH. Дослідження ортологічних кластерів генів проводили із використанням OrthoVenn.

Дипломну роботу «Дослідження кластерів генів вторинних метаболітів штаму *Streptomyces ambofaciens* Myt7 ізольованого з Чорного моря» викладено на 75 сторінках друкованого тексту, вона включає 31 рисунки та 9 таблиць. Наведено посилання на 80 джерел літератури (2 кирилицею та 78 латиницею).

Ключові слова: *Streptomyces ambofaciens* Myt7, кластери генів, вторинні метаболіти, біоінформаційні методи.

Bioinformatic analysis of *Streptomyces ambofaciens* Myt7 genome isolated from the Black Sea was carried out to study gene clusters of secondary metabolites and clarify the prospects for the use of these secondary metabolites. Research of the presence of gene clusters was carried out using antiSMASH program. Research of orthologous gene clusters was performed by using OrthoVenn.

Diploma thesis “Study of gene clusters of secondary metabolites of *Streptomyces ambofaciens* Myt7 strain isolated from the Black Sea” is expounded on 75 pages, it contains 31 figures and 9 tables. It provides links to 80 references (2 cyrillic and 78 latin).

Key words: *Streptomyces ambofaciens* Myt7, gene clusters, secondary metabolites, bioinformatics methods.

ЗМІСТ

ВСТУП	4
1. ОГЛЯД ЛІТЕРАТУРИ	6
1.1. Актиноміцети: загальна характеристика та класифікація	6
1.2. Вторинні метаболіти стрептоміцетів	10
1.2.1. Полікетидсинтази PKS	12
1.2.2. Нерибосомальні пептидсинтетази	17
1.3. Вторинні метаболіти <i>Streptomyces ambofaciens</i>	21
2. МАТЕРІАЛИ І МЕТОДИ ДОСЛІДЖЕНЬ	22
2.1. Біоінформатичний аналіз штаму <i>Streptomyces ambofaciens</i> Myt7	22
2.2. Дослідження ступеню ідентичності між отриманими штамми	22
2.3. Філогенетичний аналіз	23
2.6. Пошук кластерів генів з використанням antiSMASH	27
2.7. Порівняння та аналіз повногеномних ортологічних кластерів	29
3. РЕЗУЛЬТАТИ ДОСЛІДЖЕНЬ ТА ЇХ ОБГОВОРЕННЯ	31
3.1. Біоінформаційний аналіз штаму <i>Streptomyces ambofaciens</i> Myt7	31
3.2. Результати філогенетичного аналізу штаму <i>Streptomyces ambofaciens</i> Myt7	35
3.3. Визначення ступеня ідентичності між отриманими штамми	40
3.4. Пошук генів антибіотикорезистентності у геномі <i>Streptomyces ambofaciens</i> Myt7	41
3.5. Ідентифікація бактеріоцинів в геномі <i>Streptomyces ambofaciens</i> Myt7	47
3.6. Аналіз кластерів генів вторинних метаболітів штаму <i>Streptomyces ambofaciens</i> Myt7	49
3.7. Пошук ортологічних кластерів генів за допомогою OrthoVenn	56
3.8. Біотехнологічна схема дослідження кластерів генів вторинних метаболітів	61
УЗАГАЛЬНЕННЯ	63
ВИСНОВКИ	65
СПИСОК ЛІТЕРАТУРИ	66

ВСТУП

Актиноміцети - грампозитивні бактерії, які здатні до міцеліального росту та утворення на повітряному міцелії гідрофобних спор. Актиноміцети є одними із найбільш цінних прокариотів в біотехнології. Вони відповідають за виробництво великої частини виявлених біоактивних речовин – вторинних метаболітів, зокрема антибіотиків, які широко використовуються в фармацевтичній промисловості. На сучасному етапі розвитку суспільства вивчення актиноміцетів набуває великого значення, адже вторинні метаболіти актиноміцетів є темою багатьох досліджень та перспективним джерелом відкриття нових фармацевтично важливих сполук [Ludwig et al., 2012].

Виникнення резистентності до антибіотиків залишається однією з головних проблем в медицині сьогодні. Отже постійно йдуть пошуки нових продуцентів активних біологічних сполук різного спектру фармакологічної діяльності задля розробки нових лікарських препаратів. Особливий інтерес привертає морська мікробіота, адже умови водного середовища забезпечують унікальні та складні біосинтетичні шляхи продукування вторинних метаболітів. Синтез вторинних метаболітів регулюється кластерами генів, зокрема типів PKS та NRPS, які кодуються в геномі стрептоміцетів. Вони є складними мультиензимними комплексами. Аналіз секвенованих геномів численних видів *Streptomyces* показав, що один вид може нести більше 30 кластерів генів вторинних метаболітів, що підтверджує ідею про те, що біосинтетичний потенціал цього роду бактерій ще далеко не повністю використаний. Вивчення структури, функцій та механізмів регуляції даних кластерів сприяє більшому розумінню практичного використання вторинних метаболітів стрептоміцетів, отже є актуальною проблемою.

Мета роботи: провести біоінформатичний аналіз геному *Streptomyces ambofaciens* Myt7 для дослідження кластерів генів вторинних метаболітів, та з'ясування перспективи використання цих вторинних метаболітів.

Для досягнення вказаної мети вирішувались наступні задачі:

1. Провести біоінформатичний аналіз геному *Streptomyces ambofaciens* Myt7

2. Методами *in silico* дослідити кластери генів вторинних метаболітів геному *Streptomyces ambofaciens* Myt7

Об'єктом дослідження є геном *Streptomyces ambofaciens* Myt7.

Предметом дослідження є кластери генів вторинних метаболітів геному *Streptomyces ambofaciens* Myt7.

1. ОГЛЯД ЛІТЕРАТУРИ

1.1. Актиноміцети: загальна характеристика та класифікація

Унікальність актиноміцетів полягає в розгалуженні тіла на гіфи у вигляді міцелію та здатності спороутворення. Розрізняють субстратний міцелій, який забезпечує актиноміцетів поживними речовинами та повітряний. Останній утворює спори на спороносних гілках, які мають різноманітну морфологію. Саме в них синтезуються вторинні метаболіти, серед яких визначено багато антибіотиків, антипухлинних сполук, пігментів та летких пахучих речовин [Пирог, 2004].

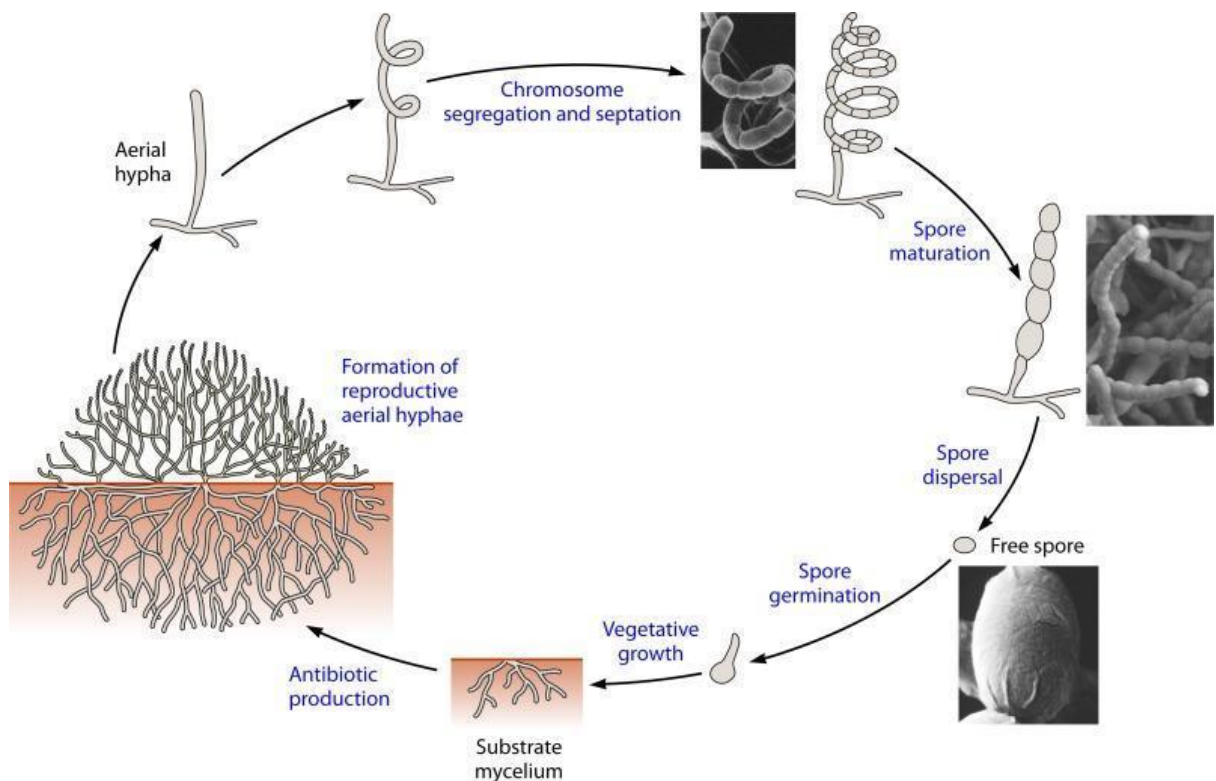


Рис. 1. Схематичне зображення життєвого циклу спорулюючих актиноміцетів [Baraka et al., 2016].

Міцеліальні або вищі актиноміцети розрізняються між собою за морфологічними властивостями: тип і стабільність міцелію, число та розташування спор, утворення спорангіїв та рухомих елементів; фізичними – теплостійкість; хімічними властивостями: будова клітинних стінок та цілих клітин, типів ліпідів [Baraka et al., 2016].

За літературними даними, представленими у другому томі бактеріального визначника Берджі, відомо, що відповідно до філогенетичної системи класифікації, створеної за даними послідовностей 16S рРНК, вищі міцеліальні та нижчі неміцеліальні актиноміцети розділяють на 9 груп:

1. Актинобактерії – включає сімейство *Actinomycetaceae*, групу кориноподібних бактерій, роди *Oerskovia*, *Promicromonospora*;

2. Актиноплани – включає актиноміцети, що утворюють спорангії та моноспоровий рід *Micromonospora*. У циклі розвитку мають рухливу стадію та стадію утворення розвиненого міцелію, розділеного перегородками. Здебільшого є сапротрофами та факультативними паразитами.

3. Актиноміцети з багатоклітинними спорангіями. У представників даної групи міцелій, що утворюється, ділиться на окремі коккоподібні клітини, рухливі у *Geodermatophilus* і *Dermatophilus* та нерухомі у *Frankia*.

4. Нокардіоподібні актиноміцети – включає роди, що містять міколові кислоти. Міцелій може розпадатися на елементи, утворюючи ланцюжки, подібні до спорангій. Справжніх спор немає. До них належать роди *Nocardia*, *Rhodococcus*.

5. Стрептоміцети – група, що поєднує різноманітні за морфологією актиноміцети, що містять діамінопімелінову кислоту у пептидоглікані клітинної стінки. Утворюють добре розвинений повітряний міцелій, що не розпадається у процесі розвитку на окремі клітини. Спорангії складаються з прямих або закручених спіраллю ланцюжків нерухомих спор, характеризуються сильною антибіотичною та хітинорозкладною активностями.

6. Мадуроміцети – конгломерат родів з різних сімейств із третім типом клітинної стінки, що містить мезо-діамінопімелінову кислоту;

7. Термомоноспори – гетерогенна група, до якої належать роди *Actinosynnema*, *Nocardiosis*, *Streptoalleteichus*, *Tliermomonospora*;

На сьогоднішній день завдяки швидкому розвитку секвенування геному багато представників актиноміцетів були перекласифіковані на основі молекулярного аналізу [Zhi et al., 2009]. Недавнім прикладом є остаточне визначення *Kitasatospora* як окремого роду всередині *Streptomycetaceae* [Girard et al., 2013]. Секвенування геному вирішило давні суперечки щодо відносин цієї групи з родом *Streptomyces* і переконливо продемонструвало, що насправді це окремий рід [Ichikawa et al., 2010; Barka et al., 2016].

В даний час новий вид не може бути заявлений без генетичного аналізу, заснованого на секвенуванні гена 16S рРНК та гібридизації ДНК-ДНК. Критерії молекулярного та хімічного складу були використані для поділу порядку *Actinomycetales* на 14 підрядів: *Actinomycineae*, *Actinopolysporineae*, *Catenulisporineae*, *Corynebacterineae*, *Frankineae*, *Glycomycineae*, *Jiangellineae*, *Jiangellineae*, *Kineosporineae*, *Microccineae*, *Kineosporineae*, *Microccineae*, *Micromonosporineae*, *Propionibacterineae*, *Pseudonocardineae*, *Streptosporangineae*. Крім того, секвенування генів 16S рРНК призвело до розпізнавання 39 сімей і 130 родів (рис. 2).

Завдяки лабільності ферментативного апарату, представники роду *Streptomyces* і роду *Micromonospora* легко пристосовуються до змін умов навколишнього середовища, а висока антагоністична активність дозволяє їм пригнічувати життєдіяльність інших бактерій. Перевага представників родів *Streptomyces* і *Micromonospora* полягає в тому, що вони здатні мінералізувати дуже стійкі та важкодоступні для інших мікроорганізмів органічні речовини, такі як хітин, целюлоза, геміцелюлоза, лігнін, водний гумус та фенольні сполуки, які потрапляють у водойми з навколишнього середовища або утворюються при відмиранні фітопланктону. Важливою також є участь актиноміцетів у деградації складних органічних сполук, таких як лігнін-целюлозний комплекс.

Целюлоза, як складова частина цього комплексу, легко руйнується під дією ендоглюконаз та екзоглюконаз актиноміцетів. Штами актиноміцетів, які мають целюлозоруйнівну активність, можуть бути рекомендовані для

біологічного очищення стічних вод целюлозо-паперової промисловості [Overmann et al., 2016].

Ранні докази існування морських актиноміцетів були отримані з опису *Rhodococcus marinonascene*, першого відкритого виду морських актиноміцетів. Подальший розвиток надійшов від відкриття, що деякі штами демонструють специфічні морські адаптації, тоді як інші виявляються метаболічно активними в морських відкладах. Останні дані наукових досліджень показали, що морські актиноміцети дійсно існують в океанах. До них належать представники родів *Dietzia*, *Rhodococcus*, *Streptomyces*, *Salinispora*, *Marinophilus*, *Solwaraspora*, *Salinibacterium*, *Aeromicrobium marinum* та інші. Серед них найбільш захоплюючою знахідкою є відкриття нового роду *Salinispora* (раніше відомий як *Salinospora*), і демонстрація широкого поширення популяцій цього роду в океані. Згодом штами *Salinispora* також були виділені з морської губки Великого Бар'єрного рифу *Pseudoceratina clavata*. Крім того, було відкрито, що штами *Salinispora* активно ростуть в деяких зразках морського осаду, що свідчить про те, що ці бактерії є метаболічно активними у природному водному середовищі [Kim et al., 2005].

Стрептоміцети становлять приблизно 10% бактерій, які колонізують морські органічні агрегати, їх антагоністична активність відіграє значну роль у деградації та мінералізації органічних речовин, іммобілізації мінеральних поживних речовин, фіксації азоту, покращення фізичних параметрів та довкілля природного середовища. [Chen et al., 2016]. Отже, актиноміцети є важливими компонентами морської мікробіоти.

1.2. Вторинні метаболіти стрептоміцетів

Вторинні метаболіти – органічні сполуки, що продукуються мікроорганізмами, грибами або рослинами, але не задіяні в життєвому циклі продуцентів.

Вони мають незначну молекулярну масу та різноманітну хімічну структуру. Синтезуються з первинних метаболітів, які, на відміну від вторинних, потрібні для життєдіяльності та росту організмів.

Особливістю вторинних метаболітів є наявність біологічної активності, що робить організмів-продуцентів суттєво важливим джерелом для розробки різних фармацевтичних та терапевтичних засобів.

В залежності від типу такої активності, вторинні метаболіти поділяють на декілька груп: антибіотики, імунорегулятори, антивірусні та антипухлинні сполуки, а також інгібітори деяких біохімічних процесів [Verdy, 2012].

Виділення вторинних метаболітів залежить від природи мікроорганізму, навколишнього середовища або методів культивування в лабораторіях. Також однією з причин утворення таких сполук можуть бути несприятливі умови, що підтверджує важливість вивчення і використання саме морських актиноміцетів у біотехнології.

Більшість вторинних метаболітів виділяють з актиноміцетів, мікроскопічних грибів та базидіоміцетів.

Для актиноміцетів характерна наявність специфічних кластерів генів – групи генів, розміщених в певній хромосомі, які кодують біосинтетичні шляхи продукування вторинних метаболітів.

Кожен з таких генів регулюється як окрема одиниця транскрипції, отже кластери не є оперонами [Сиволоб, 2008].

Розрізняють два основних класи кластерів генів – полікетидсинтази, які існують трьох типів: I PKS, II PKS, III PKS та нерибосомальні пептидсинтази (NRPS).

Найбільш відома інформація про полікетидсинтази наведена у таблиці 1.

Таблиця 1

**Відмінності між трьома типами PKS щодо структури
та механізму синтезу**

Тип PKS	Білкова структура	Механізм синтезу	Зустрічається
PKS типу I (модульний)	Один білок із кількома модулями .	Лінійний (стиль складальної лінії), в якому кожен активний сайт використовується лише один раз.	Бактерії
PKS типу I (ітеративний)	Один білок із одним модулем.	Ітеративний, коли активні сайти використовуються повторно.	Гриби
PKS типу II	Декілька білків, кожен з яких має один монофункціональний активний центр.	Ітеративний, в якому активні сайти можуть використовуватися лише один раз або багаторазово.	Бактерії
PKS типу III	Один білок із кількома модулями	Ітеративний, коли активні сайти повторно використовуються повторно.	Бактерії та рослини

1.2.1. Полікетидсинтази PKS

Полікетиди – це структурно різноманітні та біологічно активні вторинні метаболіти, отримані з природних джерел, здебільшого мікроорганізмів — бактерій, деяких рослин та грибів. Ці сполуки широко використовуються у фармацевтичній та медичній біотехнології як лікарські засоби проти різних гострих і хронічних захворювань. Полікетиди володіють різноманітними біологічно активними характеристиками, такими як протиракові, антихолестеринові та протимікробні властивості. Прикладами таких вторинних метаболітів є антибактеріальні (еритроміцин і тетрациклін), протипухлинні (антрациклін і доксорубіцин), протигрибкові (амфотерицин) і антихолестеринні (ловастатин) препарати [Piel J., 2010].

Полікетиди синтезуються шляхом декарбоксилатної конденсації карбонових кислот полікетидсинтазами (PKS). PKS поділяються на три групи залежно від їх послідовностей, первинної структури та каталітичних

механізмів, включаючи PKS типу I, PKS типу II, PKS типу III та нерибосомальні пептидсинтетази (NRPS).

PKS є ферментними комплексами, що складаються з кількох доменів, розділених спейсерними послідовностями. Першим доменом є ацетилтрансферазний – АТ, який завантажує стартову одиницю ацил-коензиму А (CoA) на ацильний білок-носії (ACP), таким чином ініціюючи біосинтез полікетидів. Другий домен є кетосинтазним – KS. Його функція полягає в поступовому подовженні вуглецевого ланцюга в ході декарбоксилативної конденсації Кляйзена. На цьому етапі біосинтезу відбувається реакція між активованою метиленовою та карбонільною групами з утворенням β -кетогрупи [Robbins T., 2016].

Різноманітні полікетидні структури виникають в ході додаткової модифікації β -кетогрупи. Це відбувається за рахунок дії додаткових доменів, зокрема кеторедуктазного (KR), дегідратазного (DH) та еноілредуктазного (ER) доменів, які поступово перетворюють дану групу в насичений ацильний ланцюг різних довжин [Fischbach M.A., 2010].

Завершує біосинтез полікетидів тіоестеразний – ТЕ-домен, який гідролізує або циклізує полікетидний ланцюг з ACP-домену для припинення подальшої елонгації сполук [Xu W., 2013].

Особливістю PKS типу III є використання ацил-КоА як початкового блоку і малоніл-КоА як субстрату для полікетидного синтезу (рис. 3). На відміну від цього, PKS типу I та PKS типу II використовують малоніл-С-пантетоїнільний остаток, який пов'язується з першим доменом ферментативного комплексу. Таким чином, ацил-КоА використовуються безпосередньо як субстрат для отримання полікетидних сполук. Для створення різноманітних структур полікетидів дана система містить ферменти, які конструюють гомодимери та каталізують багато реакцій, таких як праймування, розширення та циклізація [Shimizu Y., 2017].

Незважаючи на свою структурну простоту, PKS типу III продукують широкий спектр сполук, таких як халкони, пірони, акридони, флороглюцини,

стильбени та резорцинові ліпіди. В останні роки PKS типу III привернули більше уваги через їх різноманітні продукти, широке поширення, щодо простої структури та легкої генетичної маніпуляції.

Бактеріальний PKS типу III (RppA) був вперше ідентифіковано у *Streptomyces griseus*, виділеного з донних осадів, який відповідає за синтез 1,3,6,8-тетрагідроксинафталіну (THN). Для біосинтезу даного метаболіту потрібно п'ять молекул малоніл-КоА для утворення проміжної структури пентакетиду, а потім відбуваються процеси циклізації та ароматизації (рис. 5). Спонтанне окислення даної структури може призвести до флавіоліну – червоного пігменту. Даний полікетид має потенціальну протипухлинну активність, адже він розпізнає біогенні аміни, які продукуються нейроендокринними новоутвореннями [Funa N., 2005].

PhlD із *Pseudomonas fluorescens* конструює продукт флороглюцину (рис. 5B) з трьох одиниць малонілу-КоА. Було виявлено, що цей фермент має широку субстратну специфічність, оскільки він може приймати C₄–C₁₂ аліфатичні ацил-КоА і фенілацетил-КоА як ініціаторів утворення трикетиду в гептакетид α-піронів в результаті послідовної конденсації малоніл-КоА.

Гомологічне моделювання PhlD показало наявність прихованого тунелю, який тягнеться від активного сайту. Цей тунель сприяє зв'язуванню довголанцюгових ацил-КоА за рахунок динамічного збільшення обсягу порожнини активного центру. Мутагенез насичення в тунелі успішно змінив субстратну специфічність PhlD, що продемонструвало, що навіть незначні зміни в обсязі тунелю можуть вплинути на здатність PhlD приймати довголанцюгові ацил-КоА, і пояснило високу реактивність PhlD з різними субстратами [Zha et al., 2006].

DpgA з *Amycolatopsis orientalis* конденсує чотири молекули малоніл-КоА з утворенням 3,5-дигідроксифенілацетил-КоА (ДФА-КоА) у шляху біосинтезу 3,5-дигідроксифенілгліцину (Dpg) (рис. 5C). Dpg є непротеогенною амінокислотою, виявленою в сімействі антибіотиків

ванкоміцину. Будівництво Dpg здійснюється за каталізу чотирьох ферментів DpgA-D. DpgA є PKS типу III, що відповідає за першу стадію біосинтезу Dpg.

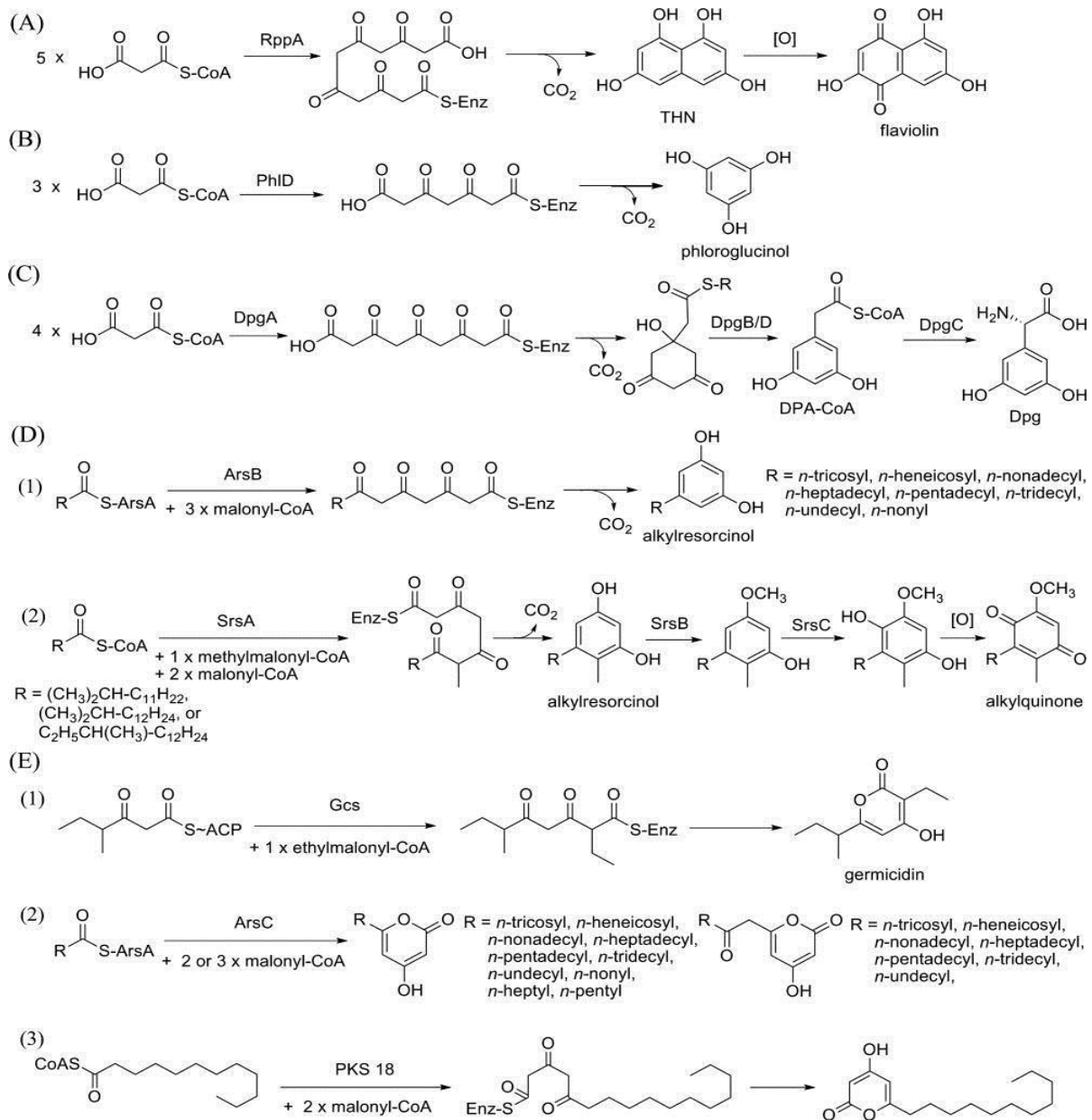


Рис. 3. Репрезентативні реакції п'яти груп бактеріальних PKS типу III.

(A) RppA каталізує утворення THN, який може бути окислений повітрям або спеціальною оксигеназою з утворенням флавіоліну.

(B) PhlD конструює продукт флороглюцину з трьох ланок малонілу-КоА.

(C) DpgA синтезує DPA-КоА із чотирьох молекул малоніл-КоА.

(D) 1. ArsB використовує три ланки малоніл-КоА для отримання алкілрезорцинів. 2. SrsA каталізує синтез алкілрезорцинів з ефірами КоА жирних кислот з розгалуженим ланцюгом як вихідні субстрати і метилмалоніл-КоА і малоніл-КоА як субстрати-подовжувачі.

(E) 1. Gcs використовує як CoA-, так і ACP-прив'язані стартери для збирання герміцидину. 2. ArsC синтезує трикетидні та тетракетидні пірони з жирних ацильних стартерів та малоніл-КоА. 3. PKS18 використовує ацил-КоА з довгим ланцюгом як стартові ланки для синтезу α -піронів у пробіріці [Dayu et al., 2012].

Гетерологічна експресія касети гена *dpgABCD* у *Streptomyces lividans* призвела до отримання 3,5-дигідроксифенілглюксілової кислоти. Це вказує на те, що три продукти гена *DpgB-D* необхідні для перетворення продукту реакції *DpgA* в амінокислоту *Dpg*. Ферментативний аналіз білка, що експресується в *S. lividans*, показав, що *DpgA* використовує малоніл-КоА як єдиний субстрат для синтезу *DPA*-КоА [Tseng et al., 2004].

Четверта група, що складається з алкілрезорцин синтаз (рис. 5D), представлена *ArsB* з *Azotobacter vinelandii* та резорцин синтазою *A* *Streptomyces* (*SrsA*) з *S. griseus*.

На додаток до двох генів ПКС III типу *arsB* і *arsC* *ars* оперон також містить два гени, які кодують два передбачуваних FAS типу I, *ArsA* і *ArsD*, відповідно. Ці два ферменти синтезують жирні кислоти з довгим ланцюгом, які безпосередньо передаються з форми, пов'язаної з АСР, активний центр *Cys* *ArsB*. Потім *ArsB* використовує жирні ацил-*ArsAs* як стартові субстрати і виконує три стадії подовження з малоніл-КоА для отримання алкілрезорцинів. Це перша демонстрація того, що FAS типу I безпосередньо взаємодіють із PKS типу III за допомогою перенесення субстрату [Miyanaga et al., 2008]. *SrsA* з *S. griseus* також може каталізувати утворення алкілрезорцинів подібно *ArsB*. Ця синтаза дуже цікава, тому що вона використовує два різні подовжувачі ланцюга, метилмалоніл-КоА та малоніл-КоА, у строго контрольованому порядку [Funabashi et al., 2008].

Остання група складається з алкілпірон синтаз (рис. 5E), які включає *ArsC* з *A. vinelandii*, герміцидин синтазу (*Gcs*) з *S. coelicolor* і нещодавно виявлену PKS18. Всі ці синтази пов'язані з побудовою різних алкілпіронов. *ArsC* приблизно на 71% ідентична *ArsB*, але на відміну від *ArsB* (яка синтезує алкілрезорцини), *ArsC* використовує жирні ацил-*ArsAs* як стартові субстрати і каталізує дві стадії подовження з малоніл-КоА з утворенням алкілпірів. *Gcs*, що кодується відкритою рамкою зчитування (ORF) *sco7221* *S. coelicolor*, бере участь у біосинтезі герміцидину і, мабуть, каталізує як подовження β -кетואцил-АСР за допомогою етилмалоніл- або метилмалоніл-КоА.

Унікальним моментом у цій реакції є те, що Gcs використовує пов'язані з АСР стартові одиниці з етилмалоніл-КоА як подовжуючу одиницю, що безпрецедентно для будь-якого іншого PKS типу III [Dayu et al., 2012].

До інших бактеріальних PKS типу III слід віднести PKS Ken2 типу III, виділений зі *Streptomyces violaceoruber*, що бере участь у продукції 3,5-дигідроксифенілглицина (3,5-DHPG). Ця сполука є непротеїногенною амінокислотою, необхідною для утворення кендоміцину та деяких інших глікопептидних антибіотиків, таких як балхіміцин, хлореремоміцин, а також ванкоміцин [Wenzel et al., 2008]. Дані антибіотики мають значну цитотоксичну активність щодо різних ліній пухлинних клітин [Wenzel et al., 2008].

Crz6, що кодується геном crz6 і виділений зі *Streptomyces sp.* MK730-62F2, як повідомляється, бере участь у біосинтезі капразаміцинів, продукуючи групу нових трикетідпіренів (пресульфідіцинів) [Tang et al., 2013]. Крім того, інше відкриття також передбачає, що DruA каталізує утворення алкілдігідропіронів в *Streptomyces reveromyceticus* (рис. 4) [Aizawa et al., 2014].

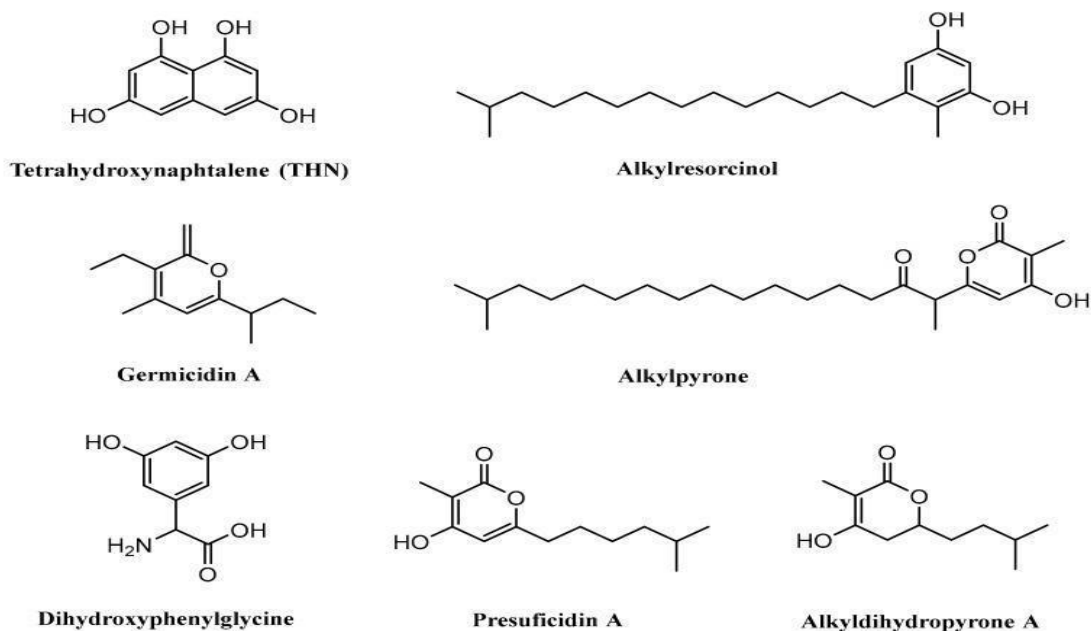


Рис. 4. Хімічна структура алкілдігідропіронів

1.2.2. Нерибосомальні пептидсинтетази

При біосинтезі вторинних метаболітів формування пептидного зв'язку може відбуватися без участі рибосом. Такі біологічно активні сполуки є нерибосомальними, а мультиензимні комплекси, які їх здійснюють, називаються нерибосомальними пептидсинтетазами (NRPS).

Можна виділити дві категорії NRPs в залежності від того, чи синтезуються вони за допомогою мультиферментного тіоматричного механізму або без нього [Sumi et al., 2015]. Перші зазвичай призводять до структур, які мають від двох до близько 50 залишків радикалів, наприклад ланцюгів жирних кислот, (тобто ліпопептиди (LP) і сідерофори), тоді як другі, як правило, менше. Продуктами NRPS є різноманітні та складні за структурою сполуки, зокрема антибіотики, імунодепресори, цитотоксини та протипухлинні агенти.

NRPS складається з модульних білків, кожний з яких містить 1000-1500 амінокислотних останків і діє як незалежний фермент. Такі блоки-модулі складаються з кількох каталітичних доменів, ковалентно пов'язаних у межах одного поліпептидного ланцюга (рис. 5). У клітині продуцента NRPS розташовані вздовж мембрани. Домени можуть мати різну геометричну форму та бути розміщені під різними кутами стосовно один одного, що обумовлює їх пластичність [Reimer, J.M., 2019].

Комплекси NRPS дуже гнучкі, адже домени з'єднані гнучкими лінкерами, які дозволяють взаємодіям між ними змінюватися під час процесу складання ланцюга. Однак окремі домени є стабільними точками вздовж каталітичного шляху. Отже вони менш динамічні і їх можна легше вивчати за допомогою таких методів, як рентгенівська кристалографія [Tagry, 2017].

NRPS можуть обробляти кілька сотень мономерів, зокрема амінокислоти, гідроксикислоти, а також гетероцикли та жирні кислоти (рис. 6). NRPS містить три основні домени, які умовно згруповані в три категорії: завантаження, елонгації та термінації. Перший з даних доменів – домен аденілювання (A). Він відповідає за вибір субстрату, утримує амінокислоту і

Другим доменом є невеликий пептидил домен білка-носія (PCP). Його функція полягає в перенесенні проміжних продуктів біосинтеза – тіоефірів до каталітичних центрів. Це відбувається за рахунок кофактора фосфопантетеїну – ФП, який ефірно пов'язаний із гідроксильною групою серина білкової частини. Джерелом ФП є кофермент А. Ця послідовність реакцій повторюється до тих пір, поки не буде утворений повнорозмірний поліпептид, тобто відбувається процес елонгації. Під час біосинтезу зростаючий пептидний ланцюг залишається ковалентно пов'язаним з 4'-фосфопантетеїнільним кофактором домену PCP, імовірно, щоб уникнути розчинення зростаючого пептидного ланцюга та забезпечити утворення правильної пептидної послідовності [Ackerley, 2016].

Завершує процес біосинтезу конденсаційний (C) домен, який каталізує утворення пептидних зв'язків між проміжними ацил-S-PCP двох сусідніх модулів. Термінаційний модуль також містить додаткову тіоестеразу. Таким чином пептидний ланцюг вивільняється з комплексу NRPS шляхом внутрішньо- або міжмолекулярної циклізації, що каталізується одним C-кінцевим доменом тіоестерази.

Модулі також можуть містити додаткові епімерази, редуктази та метилтрансферази, які забезпечують різні модифікації пептиду, наприклад, окислення, галогенування, гетероциклізацію та глікозилювання. Це ще більше розширює різноманітність хімічних структур вторинних метаболітів, що продукуються.

Структурна характеристика ключових комплексів є важливою метою покращення розуміння селективності та функціонування NRPS. Наприклад, структури доменів А в поєднанні з PCP можуть існувати в різних станах. Це відповідає зв'язуванню субстрату, його активації та завантаженню ФП та дає уявлення про механізми, що лежать в основі селективності домену А. Проте структурно охарактеризувати домени С виявилось більш складним, оскільки на сьогоднішній день повідомляється про меншу кількість прикладів такої взаємодії. Крім того, не повідомлялося про жодну структуру домену С в

комплексі з акцепторним доменом РСР, що несе субстрат. Це підтверджує наявність специфічності С домену [Mullowney et al., 2018].

1.3. Вторинні метаболіти *Streptomyces ambofaciens*

Спіраміцини I, II і III є макролідами з 16-компонентним кільцем з антибіотичною активністю, що використовуються в терапії людей як антибактеріальні та протипаразитарні засоби (активні відносно *Toxoplasma spp.*) [Chew et al., 2012; Poulet et al., 2005]. Вони містять ядро макролактонового кільця (платенолід I), до яких два аміноцукри (мікаміноза, форозамін) і один нейтральний цукор (мікароза) приєднані.

Біосинтез спіраміцину досліджувався більше ніж 30 років і генний кластер, який керує цим біосинтезом у *Streptomyces ambofaciens* – *spm* cluster. Синтезується макролактоновий каркас спіраміцину PKS, що складається з п'яти субодиниць (*spm8* до *spm12*) організований у 8 модулів і 36 доменів, причому домен кеторедуктази модуля 4, ймовірно, неактивний. Окрім класичних прекурсорів малоніл-КоА та метилмалоніл-КоА, цей PKS також містить етилмалоніл-КоА та метоксималоніл-КоА. Ферменти, що беруть участь у біосинтезі цих подовжувачів кодуються кластером *spm*: *Spm4* кодує кротоніл-КоА-редуктазу, яка, ймовірно, бере участь у синтезі етилмалоніл-КоА, і *Spm3*, *Spm34*, *Spm35*, *Spm36* і *Spm37* беруть участь у біосинтезі білка-носія метоксималонілацилу [Kim et al., 2002].

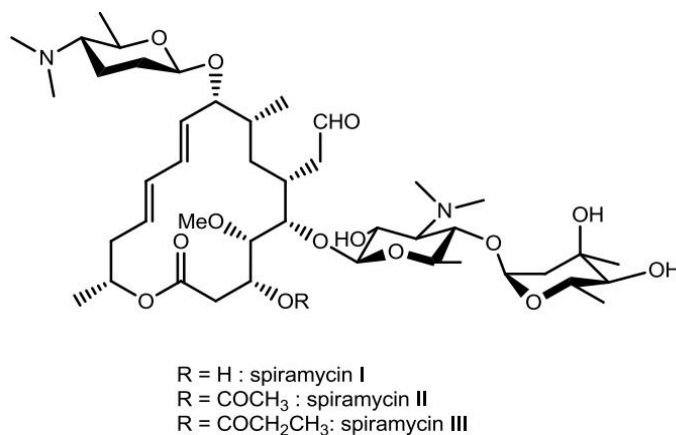


Рис. 7. Хімічна структура спіраміцину [Aigle et al., 2013].

2. МАТЕРІАЛИ І МЕТОДИ ДОСЛІДЖЕНЬ

Для проведення досліджень був використаний штам *Streptomyces ambofaciens* Myt7, виділений з поверхні мушлі мідій, ізольованих з Чорного моря, співробітником кафедри кандидатом біологічних наук Коротаєвою Надією Володимирівною.

2.1. Біоінформатичний аналіз штаму *Streptomyces ambofaciens* Myt7

Для побудови мапи генома *Streptomyces ambofaciens* Myt7 використовували сервери CGView Server (<http://cgview.ca/>) та PATRIC (<https://www.patricbrc.org/>).

Для пошуку близькоспоріднених геномів ми використовували локальний BLAST+ (версія 2.9) та встановлені відповідні бази – ref_prok_rep_genomes (1-19), сервер Type (strain) Genome Server (<https://tygs.dsmz.de/>) та antiSMASH, де обирали геноми найбільш подібні до *Streptomyces ambofaciens* Myt7.

2.2. Дослідження ступеню ідентичності між отриманими штамми

Для дослідження ступеню ідентичності між штамми використали алгоритм OrthoANI (Orthologous Average Nucleotide Identity Tool), який складається з трьох кроків (рис. 8). ANI розраховується на основі двох послідовностей генома (запитуваного і досліджуваного штамів) наступним чином – спочатку послідовність генома запитуваної штаму розділяється на послідовність (фрагменти) довжиною 1020 п.н. Потім кожен фрагмент досліджується по всій довжині досліджуваного генома розглянутого штаму з використанням програми NCBI BLASTn. У цьому процесі програма BLASTn обчислює значення ідентичності нуклеотидів між фрагментами запитуваного штаму і геному досліджуваного штаму. Середня ідентичність нуклеотидів являє собою середнє значення цієї ідентичності нуклеотидів. Пропоноване порогове значення для розмежування видів становить 95 ~ 96% як для OrthoANI, так і для вихідного ANI.

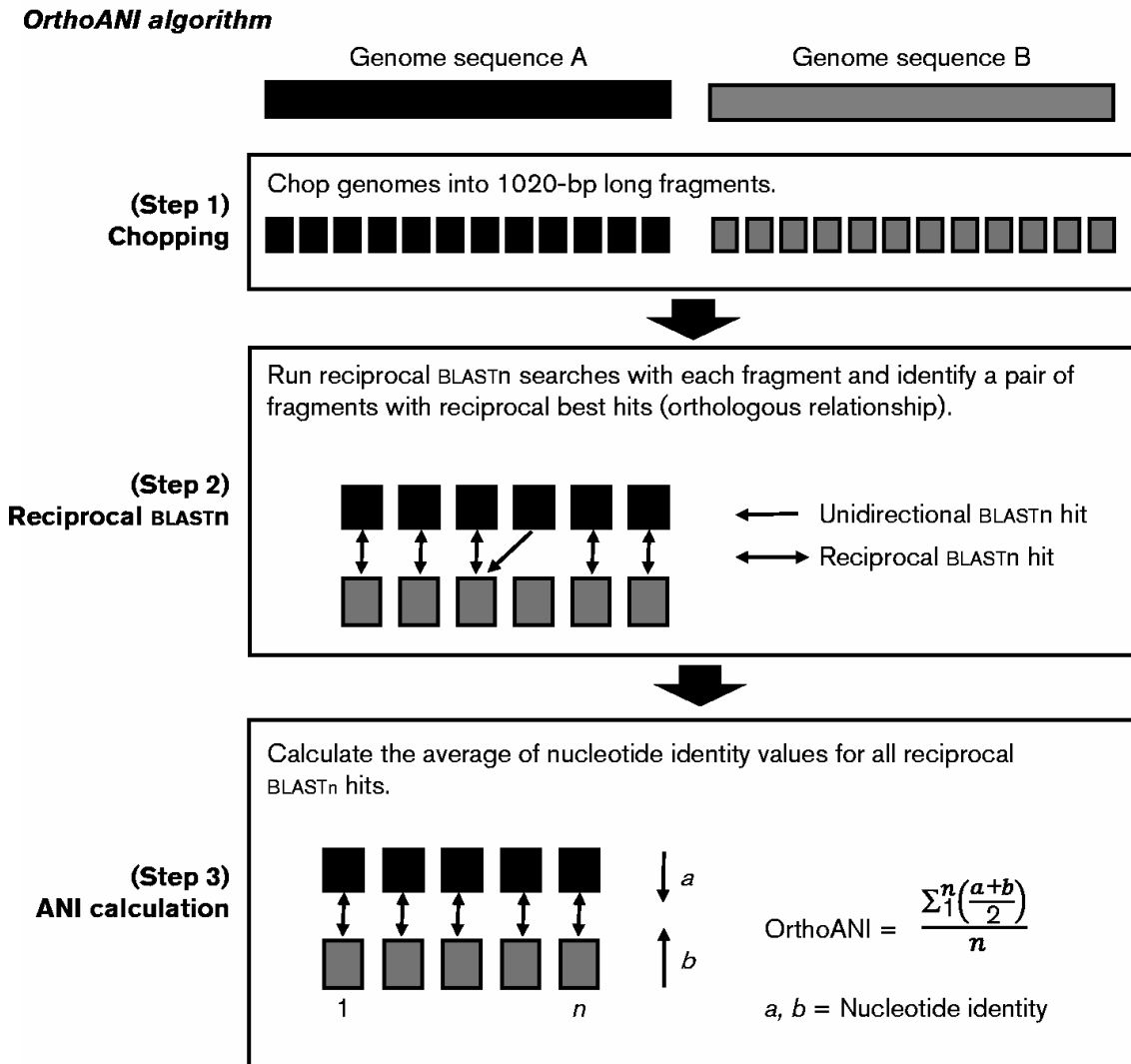


Рис. 8. Принципова схема алгоритму OrthoANI [Lee et al., 2016]

2.3. Філогенетичний аналіз

Для визначені ступеня ДНК-ДНК гібридизації (DDH) використовували сервер GGDC (<https://ggdc.dsmz.de/home.php>) та сервер Type (Strain) Genome Server (TYGS) (<https://tygs.dsmz.de/>). TYGS ґрунтується на вже існуючій інфраструктурі GGDC, але додає безліч таксономічно корисних можливостей. TYGS дозволяє замінити стандартні таксономічні методи на основі геному, такі як гібридизація ДНК:ДНК (DDH), вимірювання вмісту G+C, секвенування гена 16S рРНК та багатолокусний аналіз послідовності. Результати TYGS включають межі видів та підвидів, масштаби геному, масштаби протеома (необов'язково) та філогенетичні дерева гена 16S рРНК з підтримкою розгалуження, посилання на таксономічну літературу та посилання на депозити

типових штамів у колекціях культур. Відповідні геноми типових штамів визначаються автоматично, але можуть бути вказані користувачем. Саме з цього серверу були отримані філогенетичні відстані у вигляді файлів newick формату.

Для візуалізації отриманого дерева використовували сервер ITOL (<https://itol.embl.de/>).

2.4. Пошук генів антибіотикорезистентності

Для пошуку генів антибіотикорезистентності в геномі *Streptomyces ambofaciens* Myt7 використовували сервер CARD (<https://card.mcmaster.ca/>). Comprehensive Antibiotic Resistance Database (CARD) – основний ресурс і одночасно база даних про антибіотикорезистентність бактерій, яка забезпечує аналіз генотипу та прогнозування фенотипу на основі підібраних публікацій і послідовностей [McArthur et al., 2013; Jia et al., 2017]. Комплексна база даних резистентності до антибіотиків (CARD) виділяється своєю високою якістю моделей виявлення резистентності, які були створені вручну, та отримані на підставі експериментально підтверджених асоціацій фенотип-генотип, про які повідомляється в науковій літературі, а також завдяки співпраці з лабораторіями охорони здоров'я та клінічної мікробіології. CARD містить бази даних, специфічні для молекулярних детермінант AMR, включаючи гомологи білка, варіанти білка та рРНК, а також регуляторні мутації в системах ефлюксу [Chen et al., 2020].

Оновлення дозволили [Alcock et al., 2023] інтегрувати у CARD нову модель онтології, яка дозволила організовувати довідкові дані про послідовність AMR генів і їх мутації за базовими специфічними механізмами резистентності, з подальшими вдосконаленнями в алгоритмі ідентифікатора генів резистентності (RGI) [Alcock et al., 2023; Madden et al., 2021].

Принцип аналізу даних у наведено на рисунку 9.

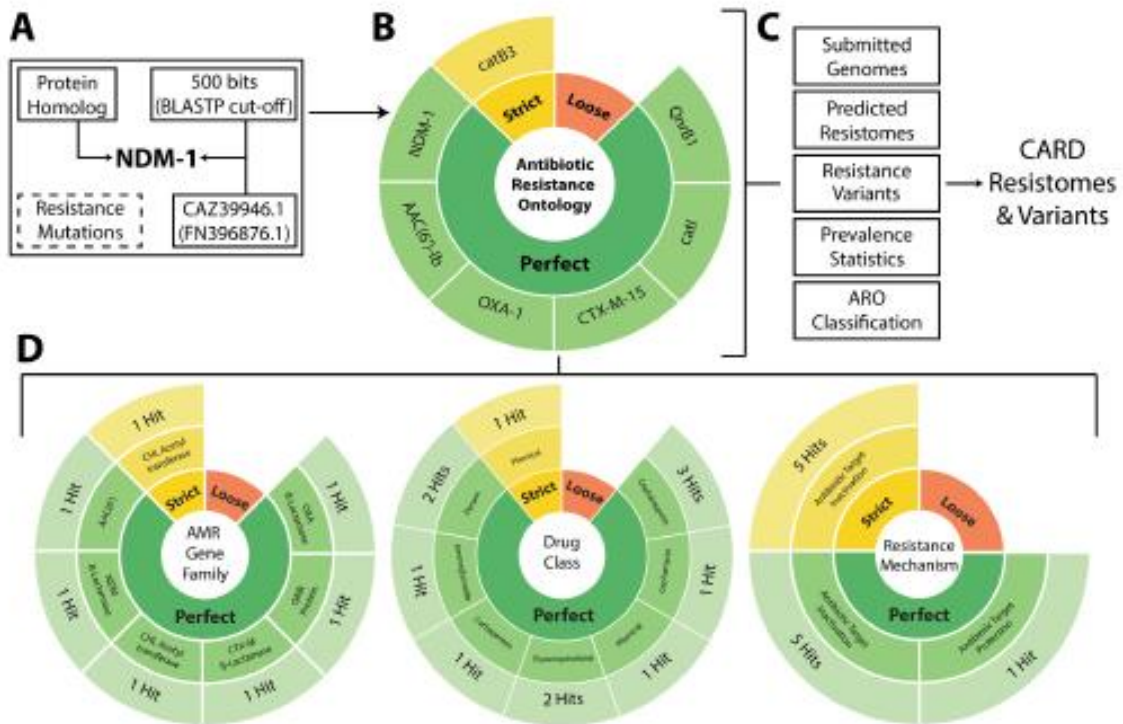


Рис. 9. Огляд програмного забезпечення Resistance Gene Identifier CARD та його роль у створенні даних CARD Resistomes & Variants. (A). Моделі виявлення AMR CARD включають еталонну послідовність, підбрану битову оцінку BLAST(P/N) і мутації, які, як відомо, передбачають AMR. (B). Запити, надіслані користувачем, аналізуються RGI за допомогою моделей виявлення, які генерують анотацію, організовану за парадигмою Perfect, Strict і Loose. (C). Під час аналізу геномних послідовностей, повних плазмідних послідовностей і збірок WGS отримані дані витягуються та обчислюються для заповнення модуля CARD Resistomes & Variants. (D). Ілюстрація тегів класифікації ARO

2.5. Пошук послідовностей RiPP та бактеріоцинових BGC

Для пошуку послідовностей RiPP та бактеріоцинових BGC в геномі *Streptomyces ambofaciens* Myt7 використовували сервер BAGEL4 (<http://bagel4.molgenrug.nl/>).

BAGEL4 (BACTERIOCIN GENOME mining tooL) один із перших інструментів, розроблених для ідентифікації RiPP та бактеріоцинових BGC. Вперше випущений у 2006 році, він був створений для вирішення проблеми, пов'язаної

з тим, що відкриті рамки зчитування (ORF) з обмеженою гомологією послідовностей важко анотувати і він працює відповідно до блок-схеми, зображеної на рисунку 10.

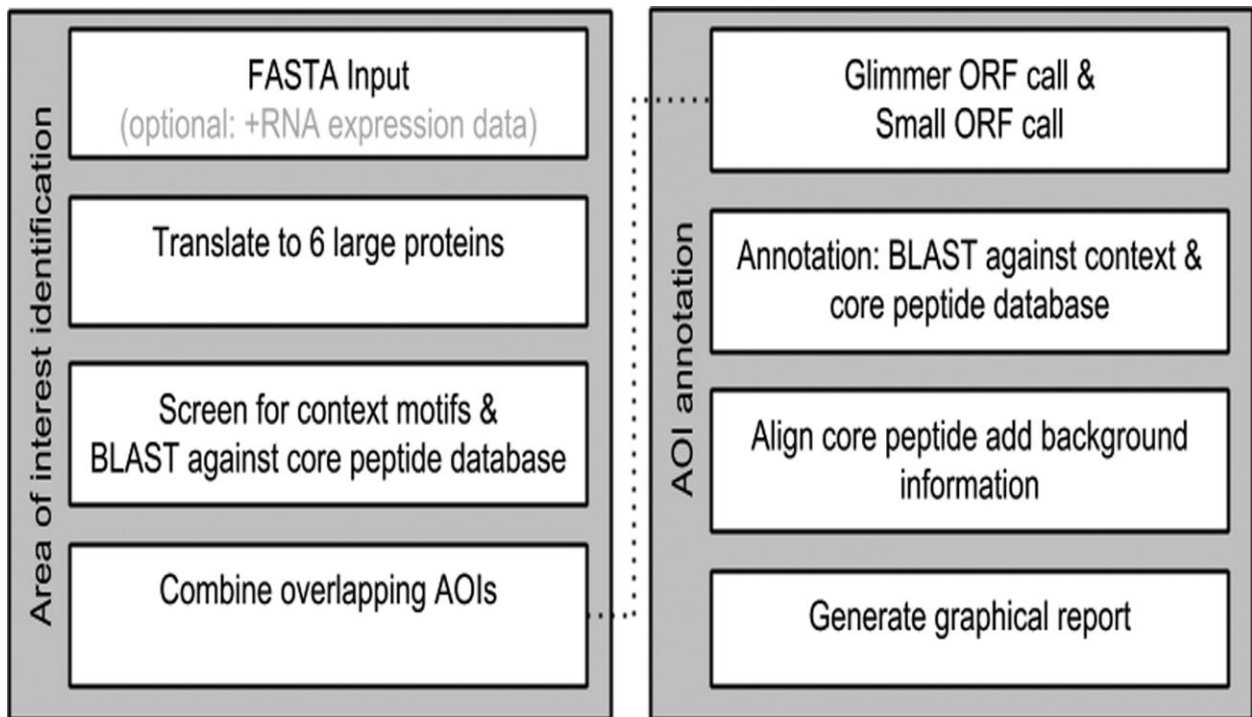


Рис. 10. Загальний потік веб-сервера VAgel4 [van Heel et al., 2018]

ДНК транлюється у шість великих білків (по одному на кожен рамок зчитування). Щоб обмежити кількість даних, які мають бути перевірені, трансляція починається лише після допустимого стартового кодону (ATG, GTG та TTG).

Згодом білки перевіряють на подібність до певних білкових мотивів і порівнюють з базою даних основних пептидів. З цих результатів вибираються так звані області інтересу (AOI). Потім об'єднуються області інтересу, що перекриваються. Всередині отриманих AOI кожен ORF шукають з використанням Glimmer3 [Delcher et al., 2007].

База даних бактеріоцинів була оновлена і тепер містить майже 500 RiPPs (клас 1), 230 немодифікованих бактеріоцинів (клас 2) та 90 великих (> 10 кД) бактеріоцинів (клас 3). Більшість записів містить посилання на NCBI або UniProt. Також, для оновлення записів використовуються спеціальні ресурси, такі як

RippMiner [Agrawal et al., 2017] та база даних MIBiG [Medema et al., 2015]. База даних доступна на сайті <http://bagel4.molgenrug.nl>.

Згодом у міжгенних областях шукають малі ORF (sORF). За замовчуванням для цих sORF мінімальна довжина складає 72 п.н. (24 амінокислотні залишки); допускається перекриття 10 п.н. з ORF, якій був знайдений за допомогою Glimmer3. Потім усі sORF порівнюють з базою даних анотацій та базою даних корових пептидів. Якщо у базі даних корових пептидів виявляється гомологія, проводиться вирівнювання. Докладний звіт складається з графічної візуалізації анотованих генів, промоторів та термінаторів.

Нарешті, створюється огляд результатів із посиланнями докладні звіти по AOI.

Пошук промотору заснований на ДНК-зв'язувальному мотиві, що розпізнається РНК-полімеразою .

Пошук термінатора реалізовано за допомогою TranTermHP [Kingsford et al., 2007] в конвеєрі BAGEL4. Прогнози візуалізуються в кластері генів і можуть допомогти зрозуміти регуляцію генів та вручну ідентифікувати основні пептиди.

Докладний звіт складається з графічної візуалізації анотованих генів, промоторів та термінаторів.

2.6. Пошук кластерів генів з використанням antiSMASH

Для кластерів генів вторинних метаболітів в геномі *Streptomyces ambofaciens* Myt7 використовували сервер antiSMASH (Antibiotics and Secondary Metabolites Analysis Shell) (<https://antismash.secondarymetabolites.org>) [Medema et al., 2011; Blin et al., 2017].

AntiSMASH є інструментом, який застосовують для пошуку та ідентифікації кластерів генів, що забезпечують продукування вторинних метаболітів мікроорганізмів. Даний онлайн сервер було запущено у 2011 році, та широко використовується наразі у багатьох біотехнологічних дослідженнях [Blin et al., 2021].

Для пошуку потенційних кластерів генів вторинних метаболітів в antiSMASH застосовуються приховані моделі Маркова (pHMM) – статистичні моделі, які ґрунтуються на аналізі та вирівнюванні множинних послідовностей сигнатурних білків та доменів, за рахунок чого, вони є ефективними для пошуку віддалених гомологів [Boddy, 2014]. Генні кластери можуть бути визначені за рахунок пошуку збігів pHMM конкретного гену, які розміщені на відстані меншій за 10 kbp. В залежності від типу, генні кластери можуть подовжуватися на ділянки довжиною від 5 до 20 kbp збоку від розміщення pHMM даного гена. Якщо кластери розташовані занадто близько, вони можуть структурно об'єднуватися та створювати гібридні моделі або перекриватися [Medema et al., 2011].

AntiSMASH найбільш точно проводить аналіз щодо структури та характеристик таких типів кластерів, що кодують NRPS, PKS та лантипептиди, які також були ідентифіковані під час даної роботи зі штамом *Streptomyces ambofaciens* Myt7.

Для здійснення аналізу геному в antiSMASH можливо використовувати як геноми з бази даних GenBank, вводячи код доступу для цього генома, так і завантажувати геном досліджуваного мікроорганізма у форматі fasta. Найновішою версією є antiSMASH 7.0, яка містить 81 тип кластерів генів, а також надає більш детальну інформацію щодо хімічної структури вторинних метаболітів та регуляції кластерів генів. antiSMASH використовує спеціальні правила, що засновані на тому, які компоненти мають бути наявними в певній області генома, щоб скласти біосинтетичний кластер генів. Цими правилами є прихована марківська модель, що якраз таки визначає послідовність даних, яка не є безпосередньо спостережуваною, але існують інші дані, що від цієї послідовності залежать. Найкраще всього antiSMASH виявляє кластери генів PKS I та II типів, NRPS, а також гібридні кластери PKS/NRPS [Blin et al., 2023].

2.7. Порівняння та аналіз повногеномних ортологічних кластерів

Порівняння та аналіз повногеномних ортологічних кластерів виконано з використанням OrthoVenn. Це потужна веб-платформа для порівняння та аналізу повногеномних ортологічних кластерів.

Гомологічні гени можна розділити на два основні класи: ортологи та паралоги. Ортологічні гени походять від загального предка під час подій специфікації [Xu et al., 2019] і зазвичай є синтетичними між близькими спорідненими видами. Паралоги також мають загального предка, але виникають у результаті подій дублювання послідовностей усередині виду і часто демонструють обмежену синтенію та більш дивергенцію, пов'язану з видоутворенням. Якщо ортологічні гени у декількох видів демонструють високу схожість послідовностей, існує ймовірність того, що вони можуть продовжувати виконувати подібні біологічні функції [Fang et al., 2010].

З іншого боку, ортологи з послідовностями, які демонструють велику дивергенцію від інших видів, з більшою ймовірністю виконуватимуть різні функції [Fang et al., 2010; Xu та ін., 2019]. В останні десятиліття ідентифікація ортологічних генів та встановлення ступеня подібності між ними стали двома важливими кроками у дослідженнях порівняльної геноміки для розуміння еволюції генів та геномів [Kristensen et al., 2011].

Було розроблено два основні методи ідентифікації ортологічних кластерів. Перший заснований на побудові дерев, та кілька інструментів, таких як програмне забезпечення PhyloTreePruner [Kocot et al., 2013] та TreeFam [Schreiber et al., 2014], використовують цю стратегію.

Другий метод заснований на графах, і ця методологія використовується в таких інструментах, як PanOCT [Fouts et al., 2012], OrthoFinder [Emms and Kelly, 2015], OrthoDB [Zdobnov et al., 2017], OMA [Altenhoff et al., 2018] та Roary [Page et al., 2015].

Алгоритми на основі дерева та на основі графа різняться у багатьох відношеннях, і кожен з них має свої переваги, але сучасні алгоритми на основі дерева, як правило, більш затратні у обчислювальному відношенні для побудови

великої кількості генів та геномів, що робить цей тип підходу менш ефективним для застосування, чим заснований на графах підхід для ортологічного аналізу великих наборів даних з більшими еволюційними відстанями [Kristensen et al., 2011]. З цих причин і OrthoVenn, і OrthoVenn2 використовували заснований на графах метод, який продемонстрував свою корисність під час автоматизації та обробки великих наборів даних [Kristensen et al., 2011; Xu et al., 2019].

Вхідними даними для OrthoVenn2 є білкові послідовності у форматі Fasta, а вихідними даними є інтерактивна таблиця шаблонів входжень та діаграма Венна з додатковими рівнями інформації та завантаженим контентом.

РЕЗУЛЬТАТИ ДОСЛІДЖЕНЬ ТА ЇХ ОБГОВОРЕННЯ

3.1. Біоінформаційний аналіз штаму *Streptomyces ambofaciens* Myt7

Анотація геному *Streptomyces ambofaciens* Myt7 була проведена з використанням серверу PATRIC (<https://www.bv-brc.org/>). Загальні статистики після проведення анотування геному наведені у таблиці

Таблиця 2

Загальні характеристики геному *Streptomyces ambofaciens* Myt7

	Характеристики	Статистика
Genome Statistics	Contigs	1
	Genome Length	8359197
	GC Content	71,57
	Contig L50	1
Genomic Features	CDS	7629
	tRNA	65
	repeat_region	
	rRNA	18
Protein Features	Hypothetical proteins	2138
	Proteins with functional assignments	5491
	Proteins with EC number assignments	1343
	Proteins with GO assignments	1166
	Proteins with Pathway assignments	1068
	Proteins with Subsystem assignments	1547
	Proteins with PATRIC genus-specific family (PLfam) assignments	6858
	Proteins with PATRIC cross-genus family (PGfam) assignments	6931
Specialty Genes	Virulence Factor (PATRIC_VF)	1
	Transporter (TCDB)	109
	Drug Target (DrugBank)	12
	Drug Target (TTD)	1
	Antibiotic Resistance (PATRIC)	50
	Antibiotic Resistance (CARD)	6
	Antibiotic Resistance (NDARO)	2

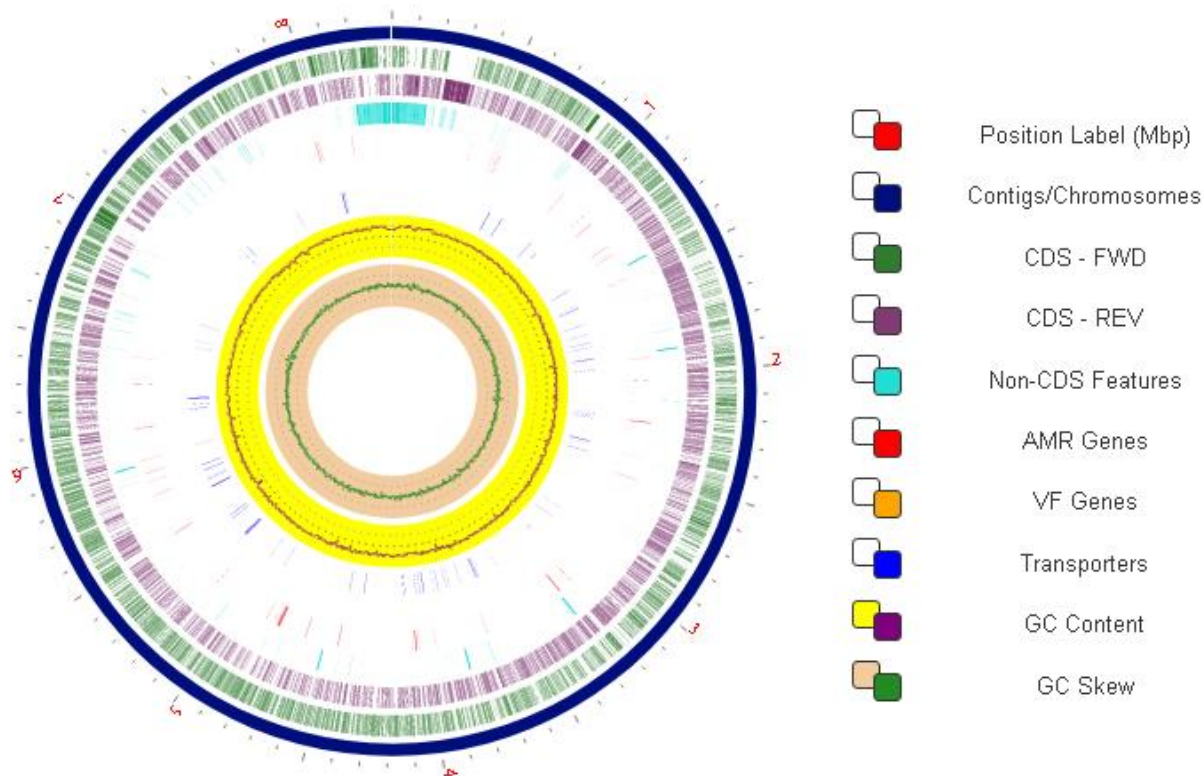


Рис. 11. Мапа, що була створена за допомоги серверу PATRIC, яка демонструє кільцеву хромосому геному *Streptomyces ambofaciens* MYT7 (оригінал)

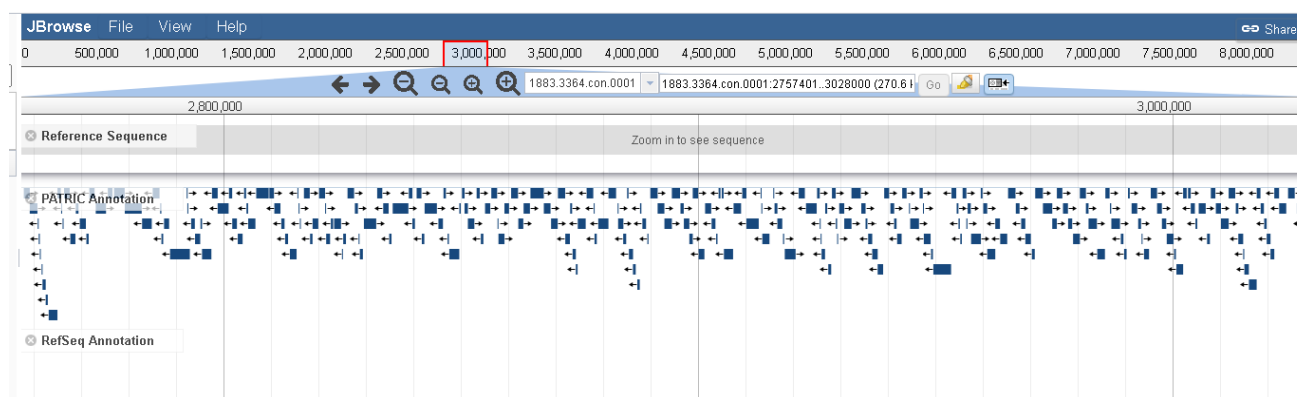


Рис. 12. Візуалізація геному *Streptomyces ambofaciens* MYT7 з використанням геномного браузеру на сервері PATRIC (оригінал)

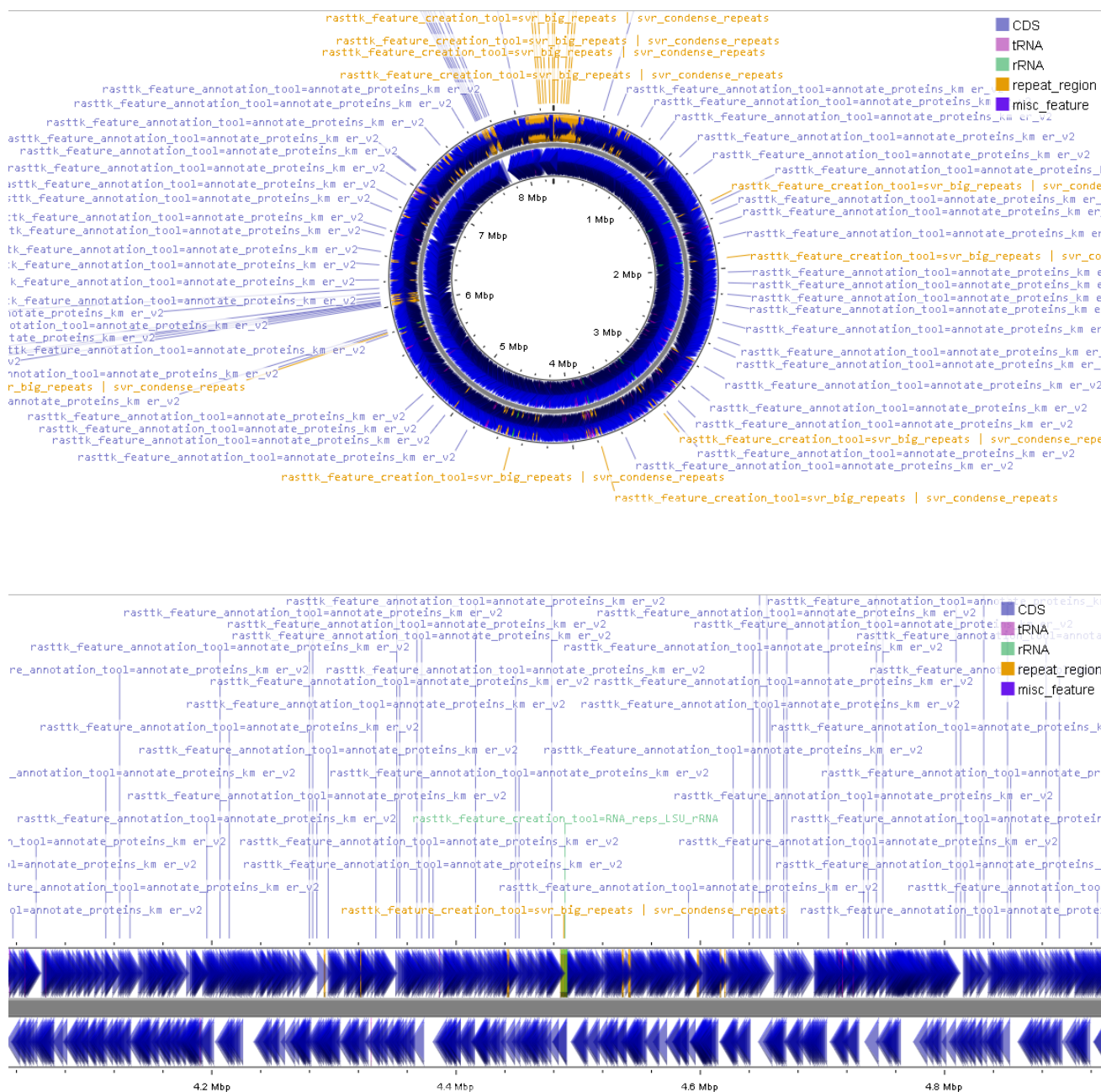


Рис. 13. Мапа, що була створена за допомоги серверу CGView, яка демонструє повний вид геному *Streptomyces ambofaciens* MYT7 та збільшене зображення (оригінал)

Для пошуку близькоспоріднених геномів були використані наступні інструменти та сервери:

- локальний BLAST+ (версія 2.9) та встановлені відповідні бази – ref_prok_rep_genomes (1-19);

- Type (strain) Genome Server (<https://tygs.dsmz.de/>);
- antiSMASH (<http://antismash.secondarymetabolites.org/>).

При використанні останнього серверу обирали ті послідовності геномів, які відносились до представників *Streptomyces*, та мали подібні кластери генів вторинних метаболітів.

Узагальнені результати наведені у таблиці 3.

Таблиця 3

**Дані про обрані геноми зі сторінки бази даних NCBI Genome
Assembly and Annotation report**

[<https://www.ncbi.nlm.nih.gov/genome/browse/#!/prokaryotes/848/>]

RefSeq	Штам	Тип фінального збирання геному	Розмір (Mb)	GC%	Джерело виділення	Географічна локація
Пошук близькоспоріднених геномів з використанням BLAST+						
NZ_CP012382.1	<i>Streptomyces ambofaciens</i> ATCC 23877	complete genome	8.3	72.2	soil	France: Picardie
Пошук близькоспоріднених геномів з використанням Type (strain) Genome Server						
BMSN00000000.1	<i>Streptomyces matensis</i> JCM 4277	whole genome shotgun sequence	7.35	72.6	soil	China
NZ_KL503831.1	<i>Streptomyces griseorubens</i> DSM 40160	whole genome shotgun sequence	7.67	72.2	soil	China: Shanghai
BMTJ01000001.1	<i>Streptomyces griseorubens</i> JCM 4383	whole genome shotgun sequence	7.67	72.2	soil	China
NZ_BMUY01000139.1	<i>Streptomyces tendae</i> JCM 4610	whole genome shotgun sequence	8.89	71.9	soil	China
NZ_BMRV00000000.1	<i>Streptomyces flaveolus</i> JCM 4032	whole genome shotgun sequence	7.87	72.2	soil	China
KN039946.1	<i>Streptomyces luteus</i> TRM 45540	whole genome shotgun sequence	7.87	72.2	soil	China: Xinjiang
BMWF00000000.1	<i>Streptomyces geysiriensis</i> JCM 4962	whole genome shotgun sequence	7.67	72.2	soil	China
NZ_BMUN00000000.1	<i>Streptomyces vinaceusdrappus</i> JCM 4529	whole genome shotgun sequence	8.46	72.5	soil	China
NZ_BMUJ00000000.1	<i>Streptomyces plicatus</i> JCM 4504	whole genome shotgun sequence	8.37	72.0	soil	China

Продовження таблиці 3

NZ_BMVQ00000000.1	<i>Streptomyces coelestis</i> JCM 4739	whole genome shotgun sequence	8.67	72.2	soil	China
NZ_BEWD00000000.1	<i>Streptomyces rubrogriseus</i> NBRC 15455	whole genome shotgun sequence	8.45	72.2	soil	China
NZ_BMPR00000000.1	<i>Streptomyces humiferus</i> JCM 3037	whole genome shotgun sequence	8.67	72.2	soil	China
Пошук близькоспоріднених геномів з використанням antiSMASH						
AL645882.2	<i>Streptomyces coelicolor</i> A3	complete genome	7.67	72.2	free-living	Europe
NZ_CP009802.1	<i>Streptomyces</i> sp. FR-008	complete genome	7.67	72.2	soil	China
NC_010572.1	<i>Streptomyces griseus</i> subsp. <i>griseus</i> NBRC 13350	complete genome	8.55	72.2	soil	China

Як ми бачимо більшість штамів не була зібрана у фінальний геном, а подана до бази даних у вигляді контігів. Для більшості актиноміцетів дійсно є складності для збирання геному у повний геном і таке явище дуже ускладнює порівняльний аналіз.

Саме тому ми не маємо можливості побудувати типове філогенетичне дерево, а будемо використовувати філогенетичні відстані, розраховані сервером TYGS. Принцип роботи цього серверу заснована на розрахунку відстаней між геномами на підставі DDH.

3.2. Результати філогенетичного аналізу штаму *Streptomyces ambofaciens* Myt7

Різниця між серверами TYGS (<https://tygs.dsmz.de/>) та GGDC (<https://ggdc.dsmz.de/home.php>) полягає в тому, що перший використовує внутрішню базу даних, а другий розраховує показник ДНК-ДНК гібридизації (DDH) між поданими геномами [Louca et al., 2019]. Алгоритми для ефективного визначення пар сегментів з високою оцінкою (HSP) або максимально унікальних збігів (MUM) добре працюють як основа для виведення міжгеномних відстаней [Auch et al., 2010 ; Meier-Kolthoff et al., 2022].

Також, зміст G+C у геномі часто використовується у таксономічних описах видів та родів. Результати показують, що вміст G+C, розрахований за послідовностями геному, варіює не більше ніж на 1% у межах виду. В даний час доцільно оцінювати вміст G + C безпосередньо з повністю секвенованих геномів. Високе охоплення, отримане за допомогою сучасних методів секвенування, забезпечує високу точність послідовності [Meier-Kolthoff et al., 2013], що, у свою чергу, вказує на те, що підрахунок цитозинів та гуанінів з геномних послідовностей є більшим точним, ніж звичайні методи.

Тому, для визначення найбільш споріднених геномів ми використали саме цей метод. Отримані результати наведені у таблиці 4.

На рисунку 14 наведено філогенетичне дерево на підставі ДНК-ДНК гібридизації (DDH), розрахованої на сервері TYGS (<https://tygs.dsmz.de/>).

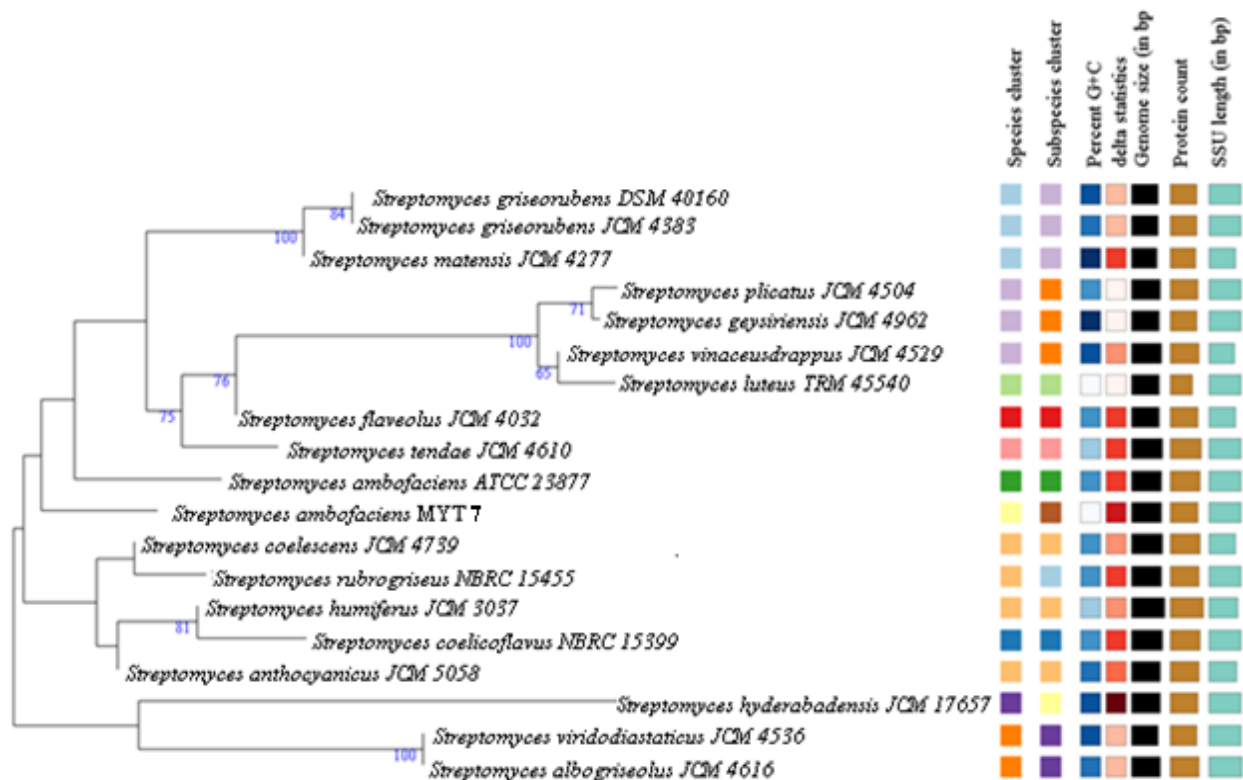


Рис. 14. Філогенетичне дерево на підставі ДНК-ДНК гібридизації (DDH), розрахованої на сервері TYGS (оригінал)

На наведеному філогенетичному дереві додатково наведена інформація про ідентичність по GC складу послідовностей, розміру геному, кількості протеїнів та довженні SSU молекули послідовності 16S rRNA.

За отриманими результатами наш штам *Streptomyces ambofaciens* MYT7 виявився максимально подібним до штаму *Streptomyces ambofaciens* ATCC 23877.

Саме цій штам за результатами аналізу з використанням BLAST+ був найбільш спорідненим до нашого штаму (ідентичність 92,0%).

Однак, отримані результати можуть бути пов'язані з тим, що інші штами не були зібрані у повний геном, що ускладнює ідентифікацію та порівняльний аналіз.

На рисунку 15 наведено окреме філогенетичне дерево, між дослідженими геномами, побудоване на підставі файлу в newick форматі:

```

((((Streptomyces ambofaciens MYT7:0.04835076,Streptomyces ambofaciens ATCC
23877:0.05383124)100:0.00264272,(((Streptomyces plicatus JCM 4504:0.00378501,Streptomyces vinaceusdrappus
JCM 4529:0.00368399)76:0.00130961,Streptomyces geysiriensis JCM
4962:0.00488389)100:0.04102644,Streptomyces luteus TRM
45540:0.04383481)100:0.01197403)100:0.00244591,(((Streptomyces hyderabadensis JCM
17657:0.03364048,Streptomyces tendae JCM 4610:0.02996952)100:0.00437926,((Streptomyces humiferus JCM
3037:0.00342596,(Streptomyces coelestis JCM 4739:0.00286666,Streptomyces anthocyanicus JCM
5058:0.00269134)71:0.00055204)86:0.01455993,Streptomyces rubrogriseus NBRC
15455:0.01640657)100:0.01843193)100:0.00493438,Streptomyces coelicoflavus NBRC
15399:0.04221928)100:0.01999869)100:0.0037319,Streptomyces flaveolus JCM
4032:0.06036386):0.0210039,((Streptomyces viridodiataticus JCM 4536:0.00709396,Streptomyces albogriseolus JCM
4616:0.00648904)99:0.02519071,((Streptomyces griseorubens JCM 4383:4.89e-05,Streptomyces griseorubens DSM
40160:3.51e-05)51:0.01068773,Streptomyces matensis JCM 4277:0.01072077)92:0.02266217)100:0.05428981);
  
```

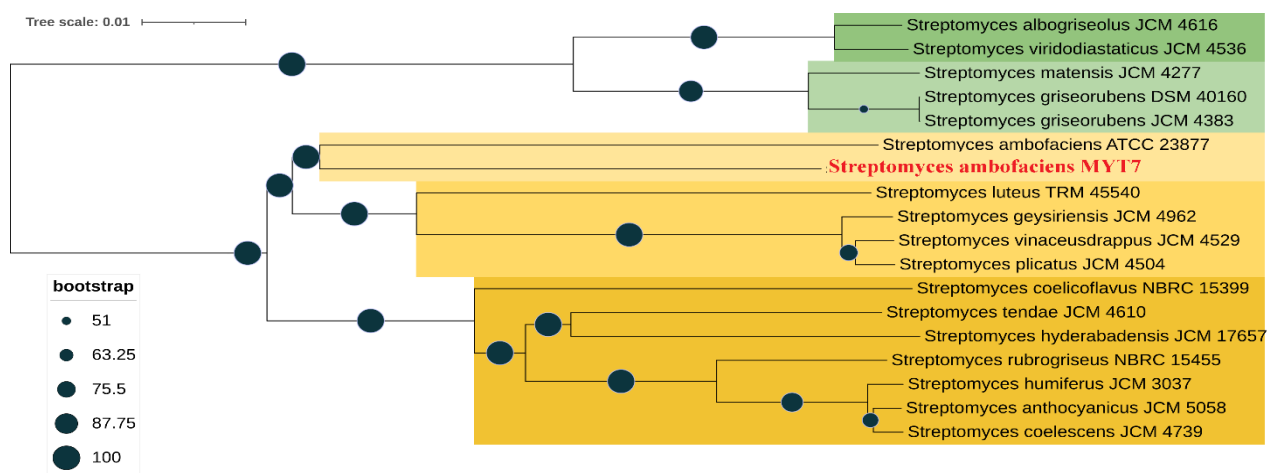


Рис. 15. Філогенетичне дерево для послідовності геному *Streptomyces ambofaciens* MYT7 та його найближчих гомологів на підставі ДНК-ДНК гібридизації (DDH), розрахованої на сервері TYGS (<https://tygs.dsmz.de/>). Візуалізація дерева виконана на сервері сервер ITOL (<https://itol.embl.de/>) (орігінал)

На рисунку 16 наведено окреме філогенетичне дерево, між дослідженими 16S rRNA, побудоване на підставі файлу в newick форматі:

```
(((((Streptomyces griseorubens DSM 40160:0,Streptomyces griseorubens JCM 4383:0)84:0.00088937,Streptomyces matensis JCM 4277:0)100:0.00288678,(((Streptomyces plicatus JCM 4504:0.00050566,Streptomyces geysiriensis JCM 4962:0.00015534)71:0.00098314,(Streptomyces vinaceusdrappus JCM 4529:0,Streptomyces luteus TRM 45540:0.00107021)65:0.0003544)100:0.00553098,Streptomyces flaveolus JCM 4032:0)76:0.00099387,Streptomyces tendae JCM 4610:0.00175964)75:0.0006452)53:0.00131913,Streptomyces ambofaciens ATCC 23877:0.00269431)22:0.0005916,Streptomyces ambofaciens MYT7:0.00212586)24:0.00032236,((Streptomyces coelestis JCM 4739:0,Streptomyces rubrogriseus NBRC 15455:0.00131104)25:0.00069125,((Streptomyces humiferus JCM 3037:0,Streptomyces coelicoflavus NBRC 15399:0.00207456)81:0.00144572,Streptomyces anthocyanicus JCM 5058:0)47:0.00039663)48:0.00132058):0.000200875,(Streptomyces hyderabadensis JCM 17657:0.00878603,(Streptomyces viridodiastaticus JCM 4536:0,Streptomyces albogriseolus JCM 4616:0)100:0.00518197)59:0.002306785);
```

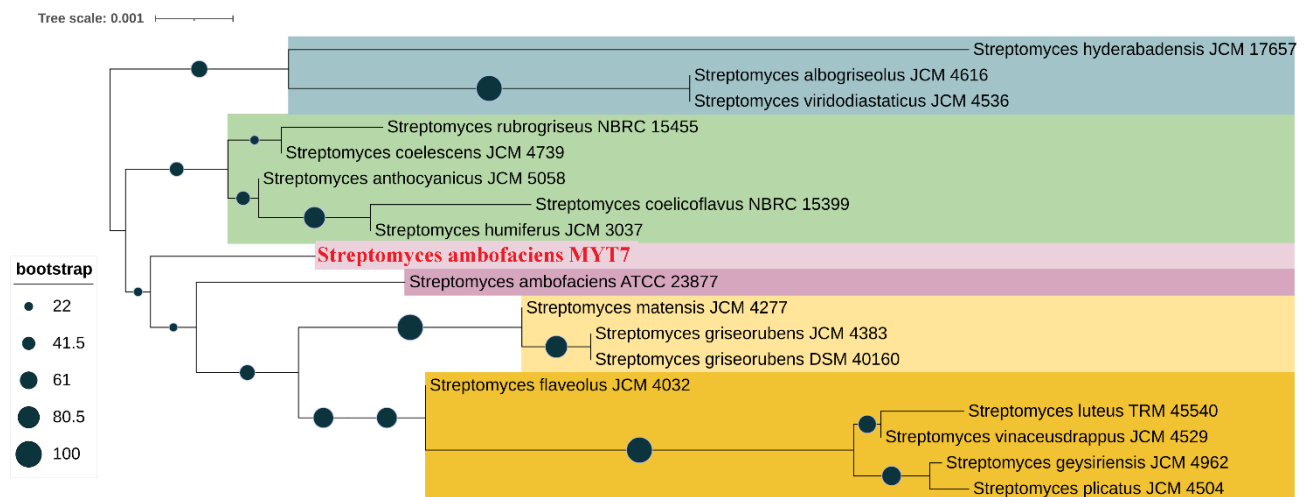


Рис. 16. Філогенетичне дерево для послідовності 16S rRNA *Streptomyces ambofaciens* MYT7 та його найближчих гомологів на підставі ДНК-ДНК гібридизації (DDH), розрахованої на сервері TYGS (<https://tygs.dsmz.de/>). Візуалізація дерева виконана на сервері сервер ITOL (<https://itol.embl.de/>) (орігінал)

Узагальнюючи усі отримані результати ми можемо сказати, що наш штам *Streptomyces ambofaciens* МУТ7 є найбільш подібним до штаму *Streptomyces ambofaciens* АТСС 23877 за існуючими даними та методами аналізу.

Показники ДНК-ДНК гібридизації (DDH) зробленої *in silico* між обраними штамми

Query genome	Reference genome	DDH	Model C.I.	Distance	Prob. DDH >= 70%	G+C difference
OD-Myt7	<i>Streptomyces_ambofaciens</i> _ATCC_23877	50.8	[47.4 - 54.2%]	0.301	13.95	
OD-Myt7	<i>Streptomyces_anthocyanicus</i> _JCM_5058	15.6	[12.7 - 19.1%]	0.8547	0	0.64
OD-Myt7	<i>Streptomyces_coelescens</i> _(anthocyanicus)_JCM_4739	15.6	[12.7 - 19%]	0.8556	0	1.33
OD-Myt7	<i>Streptomyces_coelicoflavus</i> _strain_NBRC_15399	14.1	[11.3 - 17.5%]	0.9189	0	1.19
OD-Myt7	<i>Streptomyces_coelicolor</i> _A3	46.5	[43.1 - 50%]	0.3338	6.76	0.54
OD-Myt7	<i>Streptomyces_geysiriensis</i> _JCM_4962	13.6	[10.8 - 16.9%]	0.9433	0	0.54
OD-Myt7	<i>Streptomyces_griseus</i> _subsp._griseus_NBRC_13350_(2)	15.6	[12.7 - 19.1%]	0.8547	0	1.45
OD-Myt7	<i>Streptomyces_humirefis</i> _(anthocyanicus)_strain_JCM_3037	15.2	[12.3 - 18.6%]	0.8713	0	1.33
OD-Myt7	<i>Streptomyces_hyderabadensis</i> _JCM_17657	13.2	[10.5 - 16.6%]	0.9601	0	1.58
OD-Myt7	<i>Streptomyces_plicatus</i> _JCM_4504	16.6	[13.6 - 20.1%]	0.8173	0	0.06
OD-Myt7	<i>Streptomyces_rubrogriseus</i> _strain_NBRC_15455	15.5	[12.6 - 19%]	0.8569	0	0.3
OD-Myt7	<i>Streptomyces</i> _sp._FR-008	21.4	[18.2 - 25%]	0.6761	0	0.41
OD-Myt7	<i>Streptomyces_tendae</i> _JCM_4610	13.3	[10.6 - 16.7%]	0.9549	0	1.83
OD-Myt7	<i>Streptomyces_vinaceusdrappus</i> _JCM_4529	14.3	[11.5 - 17.7%]	0.9095	0	1.04

3.3. Визначення ступеня ідентичності між отриманими штамми

Для дослідження ступеня ідентичності між штамми було використано алгоритм OrthoANI. Оскільки можливості програми обмежені 10 геномами для одночасного аналізу, були обрані геноми найбільш споріднені до *Streptomyces ambofaciens* Myt7 за результатами попередніх аналізів. Отриманий результат наведено на рисунку 17.

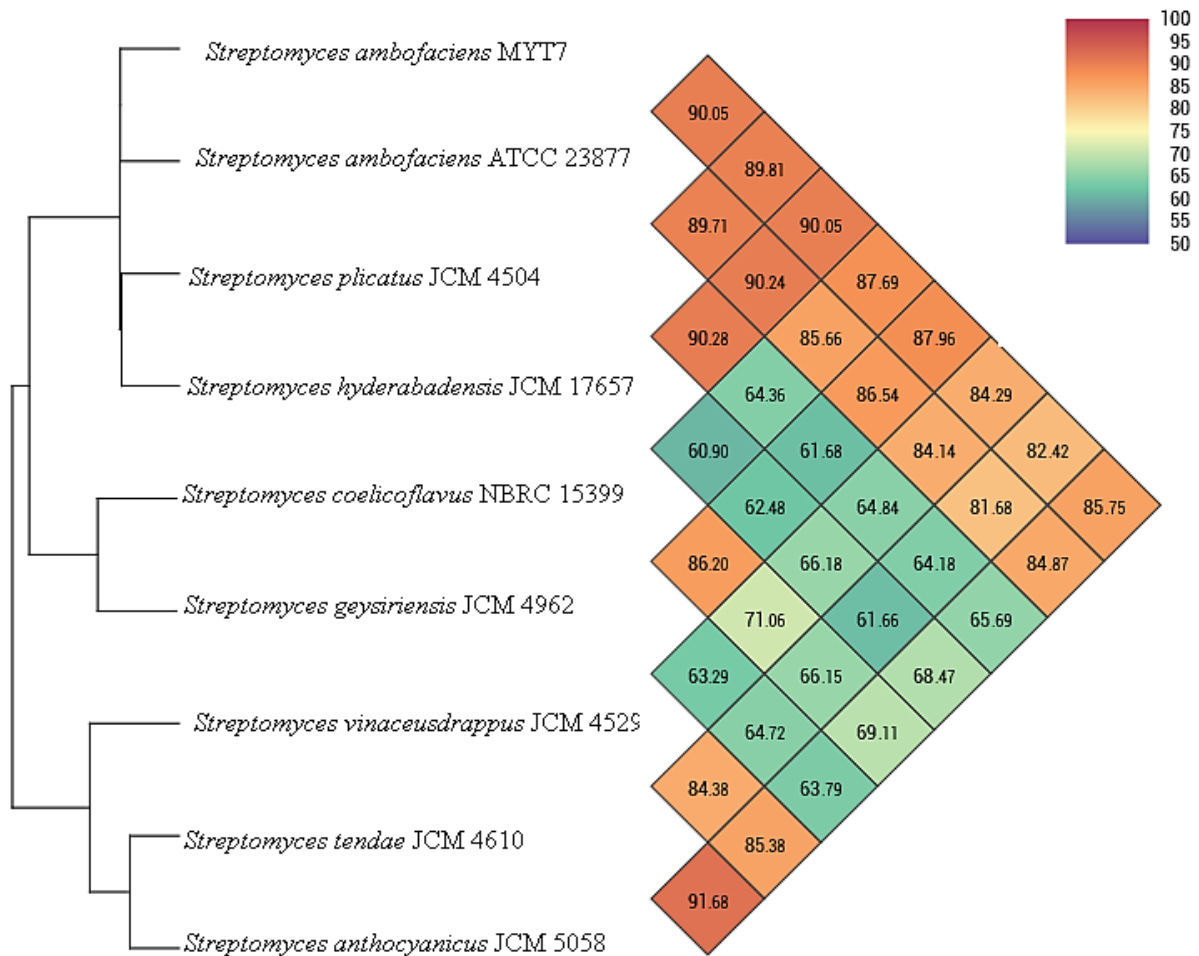


Рис. 17. Теплова карта, що показує середню ідентичність нуклеотидів за ортологією (OrthoANI) між штамом *Streptomyces ambofaciens* MYT7 та іншими спорідненими штамми, знайденими за допомогою BLAST+ та ДНК-ДНК гібридизації (DDH) і розрахована за допомогою програмного забезпечення OAT

Отримані результати також підтверджують, що найбільш спорідненим до штаму *Streptomyces ambofaciens* Myt7 виявився штам *Streptomyces ambofaciens* ATCC 23877 з показником ідентичності 90,2%.

3.4. Пошук генів антибіотикорезистентності у геномі *Streptomyces ambofaciens* Myt7

За результатами аналізу наявності генів антибіотикорезистентності штаму *Streptomyces ambofaciens* MYT7 на сервері CARD (<https://card.mcmaster.ca/>) було показано наявність лише однієї послідовності гену антибіотикорезистентності, знайденою за алгоритмом Strict. Генів антибіотикорезистентності за алгоритмами Perfect та Loose не було знайдено (рис. 18).



Рис. 18. Ідентифіковані гени антибіотикорезистентності в геномі *Streptomyces ambofaciens* MYT7 за різними алгоритмами (оригінал)

Знайдена послідовність відноситься до макролідних антибіотиків. Загальні характеристики ідентифікованої послідовності наведені в таблиці 5.

Результат ідентифікації генів антибіотикорезистентності в геномі

Streptomyces ambofaciens MYT7

Критерії пошуку	Результат пошуку
алгоритм	Strict
E-value	1.4e-218
Percent Identity	90.1
ідентифікована послідовність	mgtA MRGRGEKRLHETSRLAYIRRMTAPAHIAMFSIAAHGHVNPVLEVIRELVARGHRVTYAIP PLFAEKVAEAGAEPKLWDSTLPGPDADPEAWGTTLLDNVEPFLDDAIQALPQLVEAYEGD EPDLILHDIAAYPARVLGHRWGVVSLSPCMVAWEGYEQEVGEPMWEEPRKTERGQAYY ARFHAWLEENGITQHPDPFAGRPDRSLVLIPKALQPNADRVDKTYTFVGCQGDRTAEG DWQRPEGAEKVVVLSLGSFTKQPAFYRACVEAFAGLPGWHVVLQVGKHVDAEELGDPVE NVEVRSWVPQLAILQQADLFVTHAGAGGSQEGLATATPMIAVPQAVDQFGNADMLQGLGV ARHLPTGEATAEALRTAALALVGDPEVAGRLKEFQAQMAQEGGTRRAADLIEAELAAARA
критерій пошуку	protein homolog model
тип антибіотичної активності	macrolide antibiotic
механізм резистентності	antibiotic inactivation
антибіотики, що кодуються	erythromycin, tylosin, azithromycin, methymycin
розташування в геномі	Start 6314615 Start 6315877

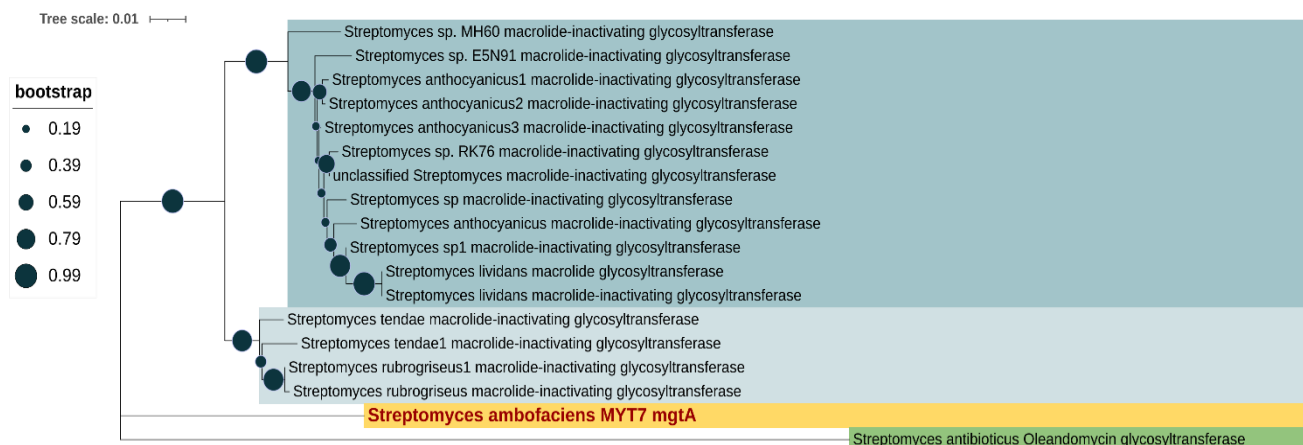


Рис. 19. Філогенетичне дерево послідовності гену *mgtA* знайденого в геномі *Streptomyces ambofaciens* MYT7. Дерево побудоване за методом Maximum Likelihood. Аналіз 15 амінокислотних послідовностей. Усі позиції, що містять прогалини та відсутні дані, були усунені. Для проведення філогенетичного аналізу використовували RAxML (Randomized Axelerated Maximum Likelihood) на сайті REALPHY з 1000 початковими повторами (оригінал)

**Гени антибіотикорезистентності, знайдені через сервер
PATRIC(<https://www.bv-brc.org/>).**

ген	Продукт	Ідентичність	E-value
cmlv	chloramphenicol resistance, MFS	83.0	1e-170
oleC	Oxytetracycline resistance,ABC-type efflux pump ATP-binding component	94.0	6e-54
novA	Efflux ABC transporter	83.0	1e-271

Для підтвердження отриманих результатів були побудовані філогенетичні дерева знайдених послідовностей з попереднім аналізом в blastp (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>).

На рисунку 20 наведено філогенетичне дерево найближчих гомологів до знайденої послідовності cmlv (chloramphenicol resistance, MFS).

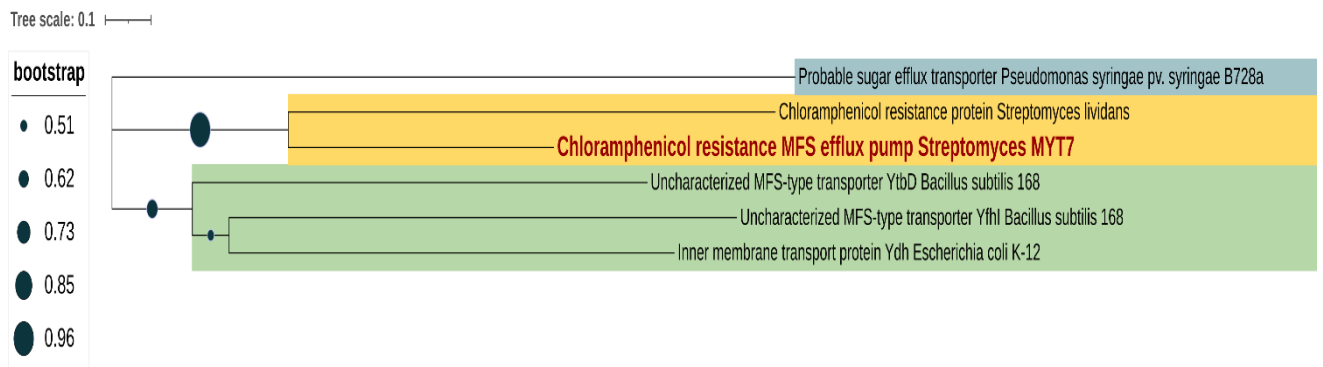


Рис. 20. Філогенетичне дерево послідовності гену cmlv знайденого в геномі *Streptomyces ambofaciens* MYT7. Дерево побудоване за методом Maximum Likelihood. Аналіз 6 амінокислотних послідовностей. Усі позиції, що містять прогалини та відсутні дані, були усунені. Для проведення філогенетичного аналізу використовували RAxML (Randomized Axelerated Maximum Likelihood) на сайті REALPHY з 1000 початковими повторами (оригінал)

Отримані результати свідчать, що в геномі *Streptomyces ambofaciens* MYT7 присутня послідовність, що є гомологічною амінокислотній послідовності гену стікості до хлорамфеніколу штаму *Streptomyces lividans*.

На рисунку 21 наведено філогенетичне дерево найближчих гомологів до знайденої послідовності *oleC* (Oxytetracycline resistance, ABC-type efflux pump ATP-binding component). Відомо, що у *Streptomyces antiteticus* існує три різні ДНК-фрагменти, що надають стійкості до олеандоміцину. Ці детермінанти стійкості до олеандоміцину позначені як *oleA*, *oleB* та *oleC*. При цьому *oleA* і *oleC* тісно пов'язані в хромосомі.

Отриманий результат підтверджує здатність штаму *Streptomyces ambofaciens* МУТ7 синтезувати вторинні метаболіти аналогічні до окситетрацикліну.

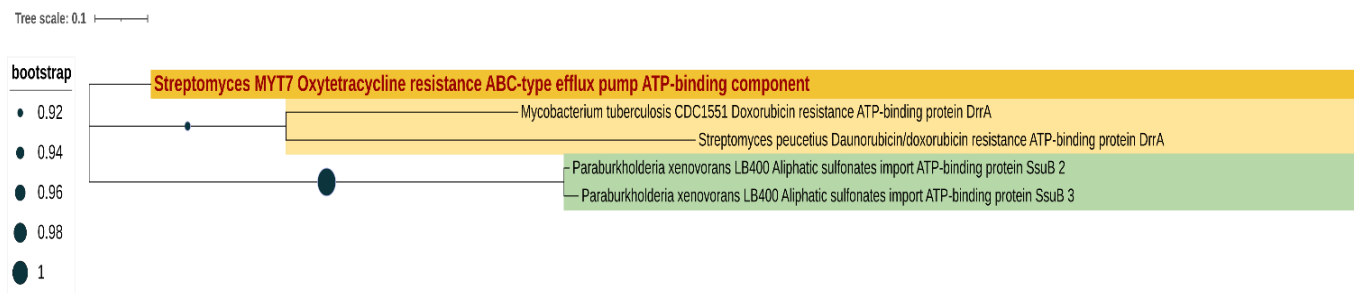


Рис. 21. Філогенетичне дерево послідовності гену *cmIv* знайденого в геномі *Streptomyces ambofaciens* МУТ7. Дерево побудоване за методом Maximum Likelihood. Аналіз 6 амінокислотних послідовностей. Усі позиції, що містять прогалини та відсутні дані, були усунені. Для проведення філогенетичного аналізу використовували RAxML (Randomized Axelerated Maximum Likelihood) на сайті REALPHY з 1000 початковими повторами (оригінал)

На рисунку 22 наведено філогенетичне дерево найближчих гомологів до знайденої послідовності *novA* (Efflux ABC transporter). Знайдена послідовність є транспортером ABC типу III, якій був ідентифікований у кластері генів біосинтезу новобіоцину, що бере участь у транспорті та стійкості новобіоцину (<https://card.mcmaster.ca/ontology/38922>). Механізм резистентності відноситься до типу ефлюкс ситеми виведення антибіотику (antibiotic efflux) [Schmutz et al., 2003]. Відомо, що *novA* і *couR5* є передбачуваними генами-транспортерами, які були виявлені в кластері новобіоцину та кумерміцину A1 у *Streptomyces coelicolor* [Schmutz et al., 2003].

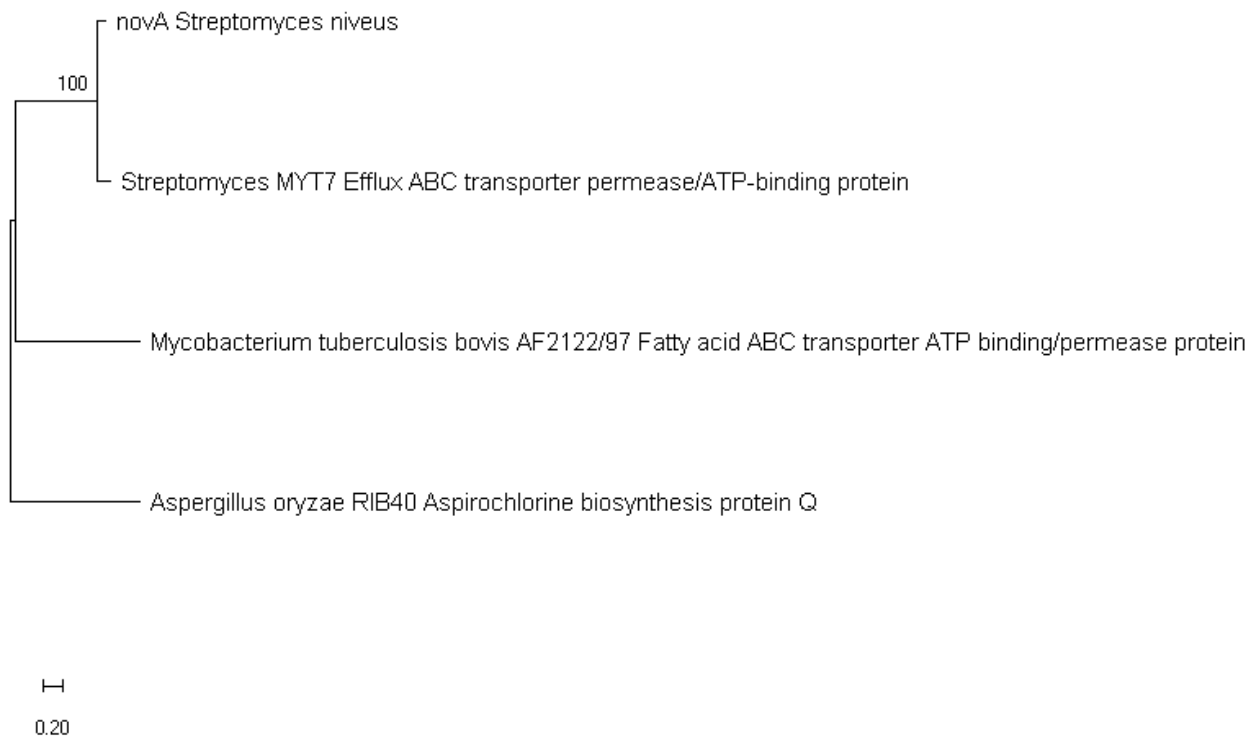


Рис. 22. Філогенетичне дерево послідовності гену *novA* знайденого в геномі *Streptomyces ambofaciens* MYT7. Дерево побудоване за методом Maximum Likelihood. Аналіз 7 амінокислотних послідовностей. Усі позиції, що містять прогалини та відсутні дані, були усунені. Для проведення філогенетичного аналізу використовували MEGAX з використанням множинного вирівнювання CLUSTAL W, та методу Neighbor-Joining з 1000 початковими повторами (оригінал)

3.5. Ідентифікація бактеріоцинів в геномі *Streptomyces ambofaciens* Myt7

З використанням веб-серверу BAGEL4 (<http://bagel.molgenrug.nl/>), було ідентифіковано послідовність, яка була ідентифікована як морфогенетичний пептид SapB. Цей пептид важливий для утворення повітряного міцелію актиноміцетів при вирощуванні клітин у багатому середовищі [Capstick et al., 2007]. Більш того, відомо, що SapB є пептидом, подібним до лантибіотиків, який отриман шляхом посттрансляційної модифікації продукту гена (*ramS*). Можливо SapB походить від RamS шляхом протеолітичного розщеплення та введення чотирьох залишків дегідроаланіну та двох лантіонінових містків [de Jong et al., 2012].

Структура кластеру наведена на рисунку 23. Гени, які входять до визначеного кластеру, наведені на рисунку 24.

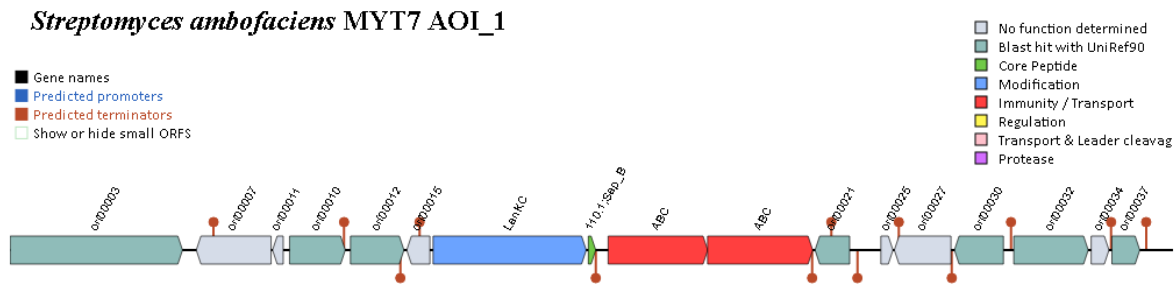


Рис. 23. Структура кластеру бактеріоцину AOI_1 (орігінал)

Condensed Gene Table

Name	Function	Motifs
orf00003	Cyclodisomaltooligosaccharide glucanotransferase OS=Bacillus circulans OX=1397 PE=1 SV=1	
orf00007		
orf00011		
orf00010	Uncharacterized oxidoreductase yghA OS=Escherichia coli O157:H7 OX=83334 GN=yghA PE=3 SV=1	
orf00012	NAD-dependent protein deacetylase 1 OS=Streptomyces coelicolor (strain ATCC BAA-471 / A3(2) / M145) OX=100226 GN=cobB1 PE=3 SV=1	
orf00015		
LanKC		lanI;labKC;PF00069
110.1;Sap_B	110.1;Sap_B	
ABC	ABC transporter ATP-binding protein RamA OS=Streptomyces coelicolor (strain ATCC BAA-471 / A3(2) / M145) OX=100226 GN=ramA PE=2 SV=1	PF00005
ABC	ABC transporter ATP-binding protein RamB OS=Streptomyces coelicolor (strain ATCC BAA-471 / A3(2) / M145) OX=100226 GN=ramB PE=2 SV=1	PF00005
orf00021	Response regulator RamR OS=Streptomyces coelicolor (strain ATCC BAA-471 / A3(2) / M145) OX=100226 GN=ramR PE=1 SV=1	
orf00025		
orf00027		
orf00030	Uncharacterized protein Rv1364c OS=Mycobacterium tuberculosis (strain ATCC Z5618 / H37Rv) OX=83332 GN=Rv1364c PE=1 SV=1	
orf00032	Vibriobactin utilization protein ViuB OS=Vibrio cholerae serotype O1 (strain ATCC 39541 / Classical Ogawa 395 / O395) OX=345073 GN=viuB PE=1 SV=1	
orf00034		
orf00037	Alanine-tRNA ligase OS=Neisseria gonorrhoeae (strain ATCC 700825 / FA 1090) OX=242231 GN=alaS PE=3 SV=1	

Рис. 24. Гени, які знайдені у послідовності бактеріоцину AOI_1 (орігінал)

Відповідно, до отриманих результатів було побудовано філогенетичне дерево корової послідовності з попереднім використанням програми blastp (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>).

При реконструкції філогенетичного дерева використовували програму MEGAX, множинне вирівнювання CLUSTAL W, та метод Neighbor-Joining (рис. 25). За результатами філогенетичного аналізу показано, що послідовність морфогенетичного пептиду, знайденого в геномі *Streptomyces ambofaciens* MYT7 є подібною до послідовності lanthionine-containing peptide SapB штаму *Streptomyces mutabilis*. В базі даних Uniprot ця послідовність представлена за номером A0A086MSR8_9ACTN.

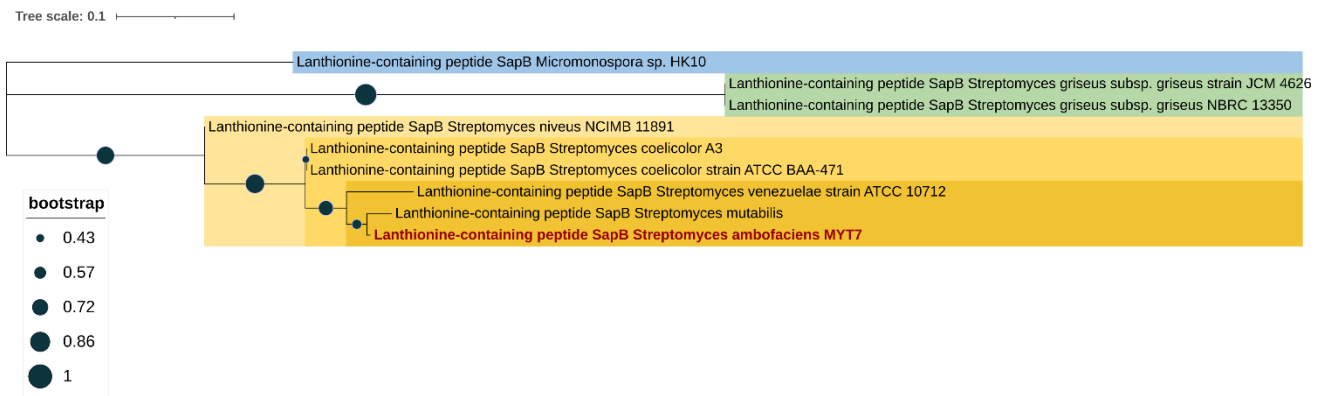


Рис. 25. Філогенетичне дерево бактеріоцину АОІ_1 (оригінал)

3.6. Аналіз кластерів генів вторинних метаболітів штаму *Streptomyces ambofaciens* Myt7

При проведенні біоінформатичного аналізу секвенованого геному штаму *Streptomyces ambofaciens* Myt7, було виявлено 34 регіони геному, пов'язаних з продукуванням вторинних метаболітів та 45 кластерів генів вторинних метаболітів.

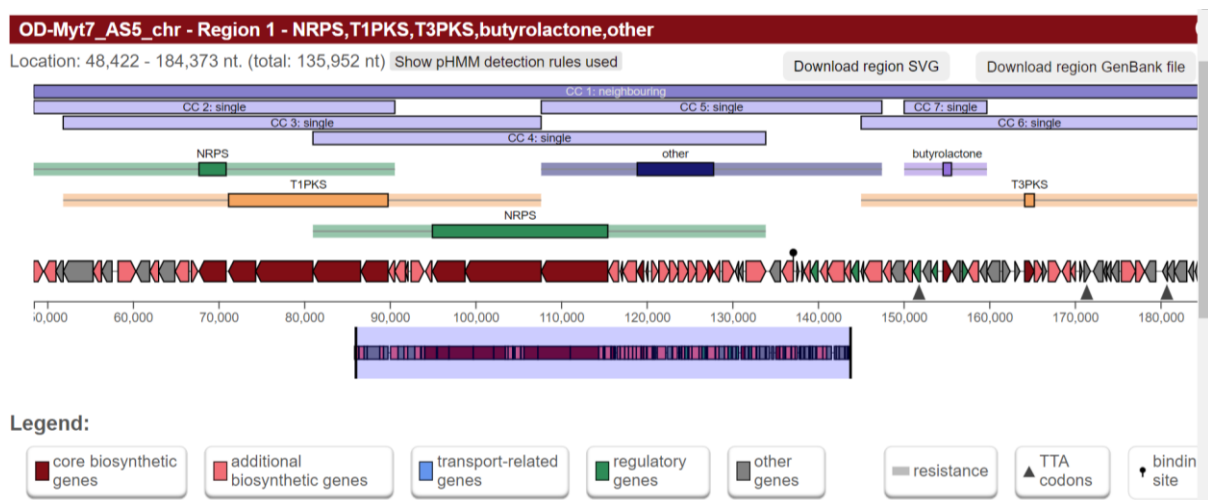


Рис. 26. Кластер-кандидат, визначений для *Streptomyces ambofaciens* Myt7 у регіоні 1

На рисунку 26 зображено формування кластеру-кандидату NRPS-T1PKS-T3PKS-butyrolactone-other у регіоні 1, що містить 5 сусідніх протокластерів. Також наведені 5 типів генів, наявні у регіоні 1 геному даного штаму. Потенціальні біосинтетичні гени зображені червоним кольором; додаткові гени біосинтезу – рожевим; гени, пов'язані з транспортом – синім кольором; гени, пов'язані з регуляцією – зеленим кольором; інші – сірим. Таким чином, основними генами

пов'язаними з продукуванням вторинних метаболітів у регіоні 1, є ctg1_53, ctg1_54, ctg1_55, ctg1_56, ctg1_57, ctg1_63, ctg1_64, ctg1_65.

Кластери генів, що продукують вторинні метаболіти у *Streptomyces ambofaciens* Myt7 та їх характеристика наведені у таблиці 7, відповідно до їх наявності у конкретному регіоні генів.

Таблиця 7

**Кластери генів, які кодують синтез вторинних метаболітів у штаму
Streptomyces ambofaciens Myt7**

Region	Type	Most similar known cluster		Similarity
1	NRPS, T1PKS, other, T3PKS, butyrolactone	polyoxypeptin	NRP+Polyketide	48%
2	T1PKS, NRPS-like	candicidin	NRP+Polyketide	95%
3	RiPP-like	Не ідентифіковано		
4	lanthipeptide-class-iii	SapB	RiPP:Lanthipeptide	100%
5	indole	5-dimethylallylindole-3-acetonitrile	Other	100%
6	terpene	isorenieratene	Terpene	75%
7	T1PKS	vicenistatin	Polyketide	75%
8	lanthipeptide-class-v	tetronasin	Polyketide	5%
9	NRPS-like	meridamycin	NRP+Polyketide	10%
10	terpene	xiamycin A	Terpene	9%
11	ectoine	ectoine	Other	100%
12	melanin	istamycin	Saccharide	4%
13	NI-siderophore	desferrioxamin B/desferrioxamine E	Other	83%
14	melanin	bagremycin A/bagremycin B	Other	44%
15	T3PKS	naphthomycin A	Polyketide	9%
16	terpene	albaflavenone	Terpene	100%
17	T2PKS	spore pigment	Polyketide	66%
18	NI-siderophore	Не ідентифіковано		
19	other, T1PKS, butyrolactone	tambjamine BE-18591	Alkaloid+Polyketide	17%

Продовження таблиці 7

20	terpene	geosmin	Terpene	100%
21	NI-siderophore	paulomycin	Other	9%
22	T2PKS	enterocin	Polyketide:Type II polyketide	90%
23	lanthipeptide-class-i	Не ідентифіковано		
24	NRPS-like	streptothricin	NRP	87%
25	lanthipeptide-class-i	borrelidin	Polyketide	5%
26	NRPS, transAT-PKS, NRPS-like	inthomycin B	Polyketide	100%
27	terpene	hopene	Terpene	100%
28	T1PKS	rifamorpholine A/rifamorpholine B/rifamorpholine C/rifamorpholine D/rifamorpholine E	Polyketide	11%
29	RiPP-like	informatipeptin	RiPP:Lanthipeptide	42%
30	phenazine	oxalomycin B	NRP+Polyketide	9%
31	NRP-metallophore, NRPS	coelichelin	NRP	100%
32	T1PKS	depsibosamycin B/depsibosamycin C/depsibosamycin D	NRP	9%
33	butyrolactone	griseoviridin/fijimycin A	NRP: Cyclic depsipeptide+Polyketide: Trans-AT type I polyketide	11%
34	other, NRPS, T1PKS	polyoxupeptin	NRP+Polyketide	48%

Після підрахунку типів кластерів генів, спостерігали наступну кількість: NRPS – 3, T1PKS – 7, T2PKS – 2, T3PKS – 2, Terpene – 5, butyrolactone – 3, NRPS-like – 4, RiPP-like – 2, lanthipeptide-class-iii – 1, lanthipeptide-class-v – 1, lanthipeptide-class-i – 2, indole – 1, ectoine – 1, melanin – 2, NI-siderophore – 3, other – 3, transAT-PKS – 1, NRP-metallophore – 1, phenazine – 1. Таким чином, найбільшу кількість склали NRPS, PKS, зокрема PKS I, та лантипептиди.

Для всіх даних кластерів були ідентифіковані продукти, окрім кластеру RiPP-like регіону 3, NI-siderophore регіону 18 та lanthipeptide-class-i регіону 23.

В двох регіонах геному – 1 і 34 спостерігали наявність NRPS та PKS I, що мали схожість 48% з ідентифікованими кластерами, які відповідають за синтез поліоксипептину – гексадепсипептиду, який має активність індукції апоптозу щодо клітин карциноми підшлункової залози людини AsPC-1 [Du et al., 2014].

У регіоні 26 кластери NRPS, transAT-PKS (PKS, в якій відсутні домени ацилтрансферази (AT) в модулях [Chen et al., 2019], NRPS-like (NRPS-подібний фрагмент) мають схожість 100% із кластерами, які продукують оксазол-трієновий антибіотик інтоміцин B [Wu et al., 2018]. У регіоні 31 розташовані кластери NRP-metallophore (кластери, які пов'язані з утворенням ліганду важких металів) та NRPS, які забезпечують синтез коеліхеліну – нерибосомального тетрапептиду сидерофору, який хелатує Fe³⁺. [Deeth et al., 2009].

Терпени – ароматичні вуглеводні компоненти, зустрічаються у 5-ти регіонах геному *Streptomyces ambofaciens Myt7*, а саме пігмент ізоренієратен у регіоні 6, потенційний противірусний препарат ксіаміцин А у регіоні 10, альбафлавенон – трициклічний сесквітерпеновий антибіотик у регіоні 16 [Moody et al., 2012], геосмін та гопен – ароматичні терпени 20-ого та 27-ого регіону відповідно.

Кандицидин є ароматичним полієновим макролідним антибіотиком, що закодований у регіоні 2 та регулюється кластером NRPS-like. Він має протигрибкову активність щодо роду *Candida* [Gil et al., 2002]. Також NRPS-like кластери представлені у регіоні 24 (87% схожості) та пов'язані з синтезом стрептотрицину – N-глікозидного антибіотику з активністю щодо грамнегативних бактерій [Morgan et al., 2023]. У регіоні 9 даний кластер потенційно забезпечує продукування мерідаміцину – неімуносупресивного макроциклічного полікетиду з нейропротекторною активністю в дофамінергічних нейронах (схожість 10%) [Jiang et al., 2011].

У регіонах 11, 12 та 14 наявні кластери, що потенційно кодують продукування осмопротектору ектоїну та пігментів меланіну (сахариду істаміцину та

антибіотиків багреміцину А і В, що виявляють помірну активність відносно грампозитивних бактерій і деяких грибів) [Bertasso et al., 2001].

У регіоні 5 розташований кластер індолу, що відповідає за синтез 5-диметилаліліндол-3-ацетонітрилу – протимікробного засобу та має протизапальну дію [Ozaki et al., 2013].

Наступний кластер – phenazine. Він наявний у регіоні 30 та може потенційно регулювати продукування оксаломіцину В (схожість 9%). Даний вторинний метаболіт є спіро- β -лактон- γ -лактамним фрагментом та має антибактеріальну, протівірусну та цитотоксичну дію [Oleksak et al., 2020].

PKS I наявна у семи регіонах геному, таким чином, він зустрічається у штаму в найбільшій кількості. У регіоні 7 даний кластер може кодувати синтез віценістатину (схожість 75%) – полікетидного антибіотику, що виявляє протипухлинну активність щодо карциноми товстої кишки людини Co-3 [Shindo et al., 1993]. У регіоні 28 PKS I пов'язаний з продукуванням полікетидного антибіотику рифаморфоліну типів А-Е, які мають активність щодо метицилінрезистентного *Staphylococcus aureus* (схожість 11%) [Xiao et al., 2017]. У регіоні 32 даний кластер може кодувати синтез антибіотику депсизобзаміцину типів В, С, D. У регіоні 19 наявні PKS I та бутиролактон, які пов'язані з продукуванням полікетидного алкалоїду тамбжаміну BE-18591, що здатен пригнічувати імунпроліферацію і гастрит [Pinkerton et al., 2010].

PKS II у регіоні 17 кодує синтез полікетиду – пігменту спор, а у регіоні 22 – бактеріоцину ентероміцину – полікетиду II типу.

PKS III наявна у регіоні 15 та відповідає за потенціальне продукування нафтоміцину А – жовтого пігменту, що має антибактеріальну, протигрибкову та протипухлинну дію [Kang et al., 2012]. Також даний кластер задіяний у регіоні 1 при синтезі поліоксипептину.

Кластери бутиролактонів також наявні у регіоні 1, що забезпечує синтез поліоксипептину та у регіоні 33 (схожість 11%), що регулює продукування антибіотику грізеовілідін, який активний проти *Brachyspira hyodysenteriae* – бактерії, що викликає дизентерію свиней [Asano et al., 2006], і депсипептидного

антибіотику фіїміцину, який має активність щодо метицилінрезистентного *S. aureus* [Sun et al., 2011].

Наступний тип кластеру – NI-сидерофори. Даний кластер пов'язаний із синтезом сидерофорів – хелатів важких металів. Він наявний у регіоні 21, який забезпечує синтез антибіотику пауломіцину, який має цитотоксичну дію проти різних клітинних ліній пухлин людини, таких як аденокарцинома підшлункової залози (MiaPaca_2), аденокарцинома молочної залози (MCF-7) і гепатоцелюлярна карцинома (HepG2). У регіоні 13 даний кластер може кодувати продукування десферіоксамінів В та Е – сидерофорів, які є гексадентатним хелатором іонів тривалентного феруму та використовується для лікування перевантаження організму залізом [Flora, 2013].

Наступну групу складають кластери лантипептидів, а саме I, III та V класу. Лантипептиди є рибосомально синтезованими (RiPPs) та посттрансляційно модифікованими поліциклічними пептидами, які здатні мати біологічну активність [Alkhalili et al., 2018]. У регіоні 4 наявний кластер lanthipeptide-class-iii, який пов'язаний із синтезом SapB, що функціонує як біосурфактант для сприяння росту повітряних гіф [Chen et al., 2018]. У регіоні 25 спостерігали кластер lanthipeptide-class-i, що регулює синтез полікетидного макроліду борелідину, що має протималарійну активність, проте клінічно не використовується через високу токсичність [Schulze et al., 2014]. У регіоні 8 наявний кластер lanthipeptide-class-v, що пов'язаний із продукуванням тетроназину – полікетід-поліефірного антибіотику, який діє як іонофор і деполяризує мембрану чутливих до препарату бактерій, здебільшого грамположитивних [Linton et al., 1994].

У регіоні 29 наявний кластер RiPP-like. Даний кластер може забезпечувати продукування рибосомально синтезованого і посттрансляційно модифікованого пептидного продукту, зокрема лантипептиду третього класу інформатипептину (схожість 42%). Біологічної активності даного пептиду не було встановлено [Liu et al., 2018].

Таким чином, більшість кластерів генів *Streptomyces ambofaciens* Myt7 здатні кодувати синтез біологічно активних вторинних метаболітів, зокрема антибіотиків, протипухлинних, антигрибкових та противірусних засобів.

Взагалі, 15 досліджених геномів, які виявились найбільш спорідненими до геному штаму *Streptomyces ambofaciens* Myt7 містили загалом 242 окремих BGC (рис. 27). Приблизно 70% усіх BGC, виявлених antiSMASH має менше ніж 50% схожості з найближчим кластерами в базі даних.

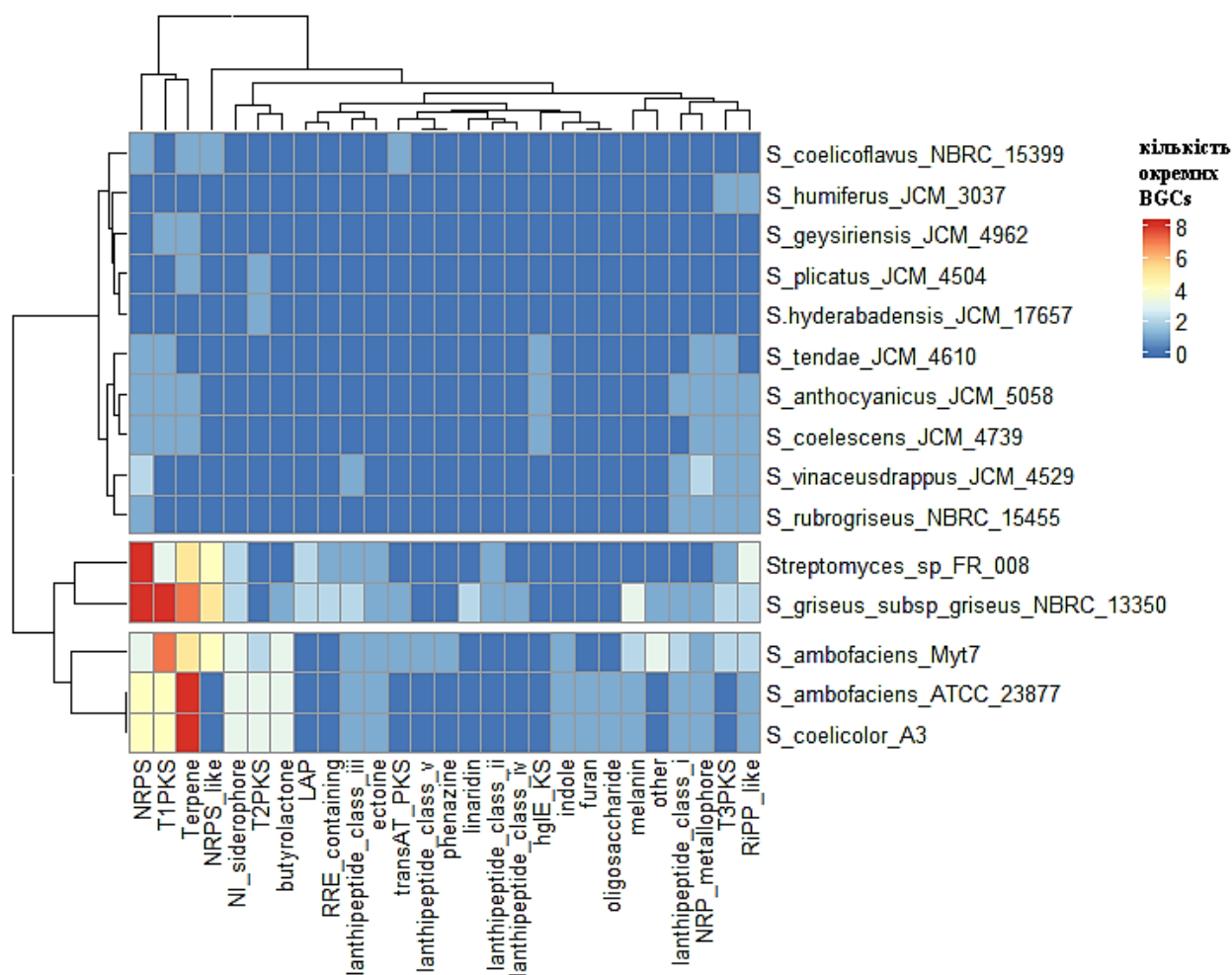


Рис. 27. Теплова мапа, яка відображає кількість окремих кластерів біосинтетичних генів, виявлених AntiSMASH в геномах досліджених бактерій (оригінал)

Порівнюючі усі проаналізовані геноми з нашим, ми з'ясували, що по кількості та організації BGCs штаму *Streptomyces ambofaciens* Myt7 відноситься до одного

угруповання зі штамми *Streptomyces ambofaciens* ATCC 23877 та *Streptomyces coelicolor* A3 (рис. 27). Однак, слід відмітити, що у штаму *Streptomyces ambofaciens* Myt7 загальна кількість кластерів була більшою (45 BGCs) в порівнянні з кількістю BGCs у штамів *Streptomyces ambofaciens* ATCC 23877 та *Streptomyces coelicolor* A3 (34 BGCs).

Так, геном штаму *Streptomyces ambofaciens* Myt7 характеризується більшою кількістю кластерів генів вторинних метаболітів наступних типів: PKS I type, PKS III type, NRPS like та RiPP like.

До наступного угруповання відносяться геноми штамів *Streptomyces sp.* FR-008 та *Streptomyces griseus* subsp. *griseus* NBRC 13350.

3.7. Пошук ортологічних кластерів генів за допомогою OrthoVenn

Для пошуку ортологічних кластерів використовували сервер OrthoVenn. Сервер має обмеження по кількості геномів, які ми маємо можливість одночасно завантажити.

Тому для аналізу використовували частину досліджених геномів. Для можливості ідентифікації відомих кластерів завантажили додатково групу *Streptomyces* з бібліотеки програми. На рисунку 28 наведена зведена таблиця зустрічальності спільних ортологічних груп серед досліджених геномів.

Цей результат кластеризації підсумовується як граф осередків, у якому кожен рядок представляє групу кластерів ортологів, а кожен стовпець вказує вид. Зелений осередок вказує на наявність кластерної групи у відповідного виду, а сіра смуга вказує на відсутність кластерної групи у цього виду.

Шаблон ліворуч вказує, які види знаходяться в кластерах, кількість кластерів – це кількість кластерів, спільних між видами, а кількість білків – це кількість білкових членів у спільних кластерах.

Гістограма з накопиченням праворуч відображає кумулятивну кількість білкових послідовностей, присутніх у кластерній групі для кожного виду.

Взагалі ми знайшли один кластер, який був присутній лише у штаму *Streptomyces ambofaciens* Myt7 (рис. 29).

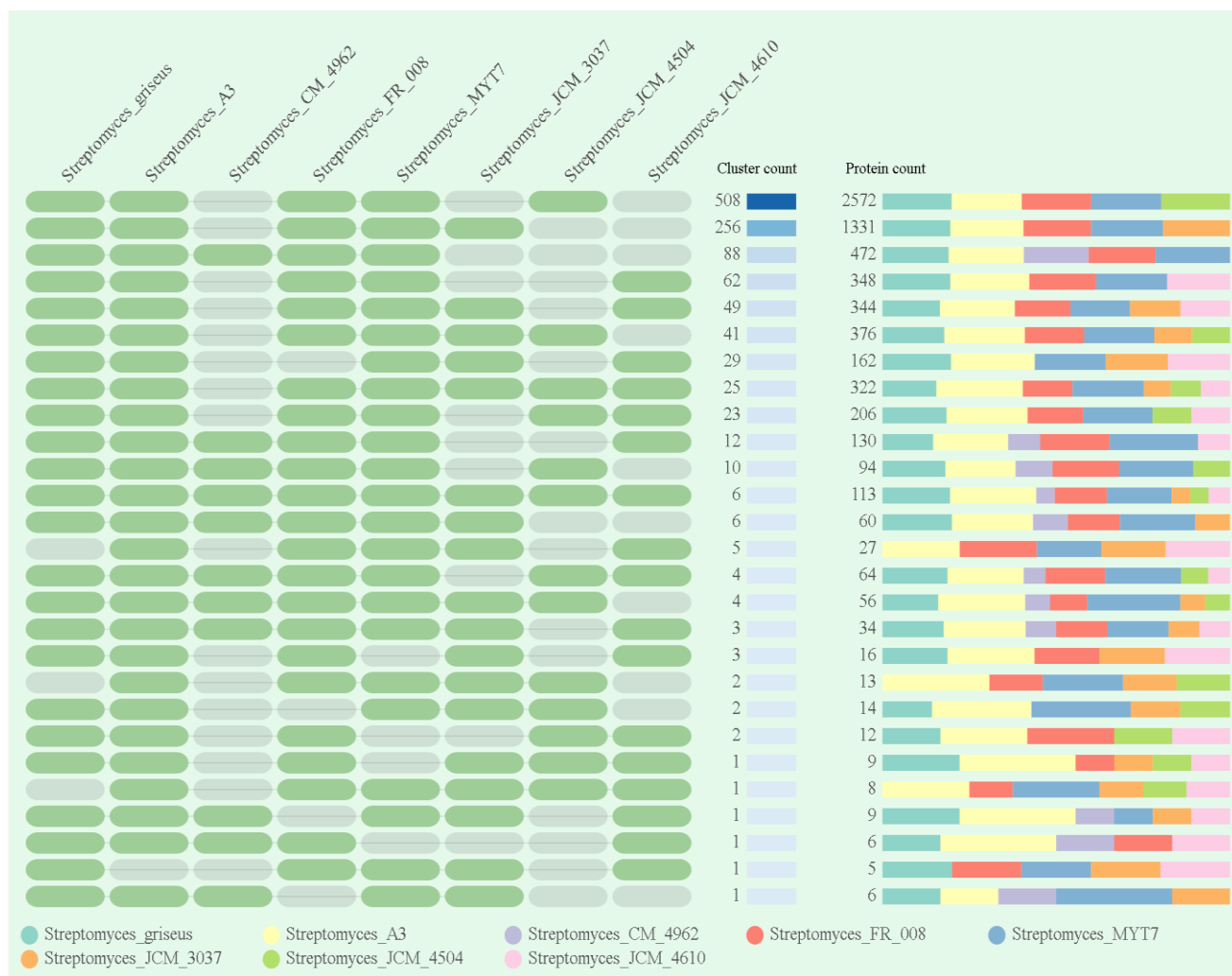


Рис. 28. Таблиця зустрічальності спільних ортологічних груп серед досліджених геномів

Слід відмітити, що в геномі штаму *Streptomyces ambofaciens* Myt7 знайдено 1378 генів, для яких не вдалося знайти ортологів в геномах інших видів (таб. 8).

Таблиця 8

Кількість білків, кластерів ортологічних груп генів та поодиноких генів

Species	Proteins	Clusters	Singletons*
Streptomyces_A3	8040	6035	1481
Streptomyces_CM_4962	209	203	6
Streptomyces_FR_008	6421	4687	1492
Streptomyces_MYT7	7629	5726	1378
Streptomyces_JCM_3037	745	727	17
Streptomyces_JCM_4504	939	873	62
Streptomyces_JCM_4610	552	497	50
Streptomyces_griseus	6930	4997	1654

Примітка: Singletons у верхньому правому куті описує гени, для яких не вдалося знайти ортологи в інших видів

На рисунку 29 наведена діаграма Венна, яка відображає розподіл спільних ортологічних кластерів серед геномів *Streptomyces ambofaciens* Myt7, *Streptomyces geysiriensis* JCM 4962, *Streptomyces* sp. FR-008, *Streptomyces coelicolor* A3, *Streptomyces humiferus* JCM 3037, *Streptomyces plicatus* JCM 4504, *Streptomyces tendae* JCM 4610 та *Streptomyces griseus* subsp. *griseus* NBRC 13350.

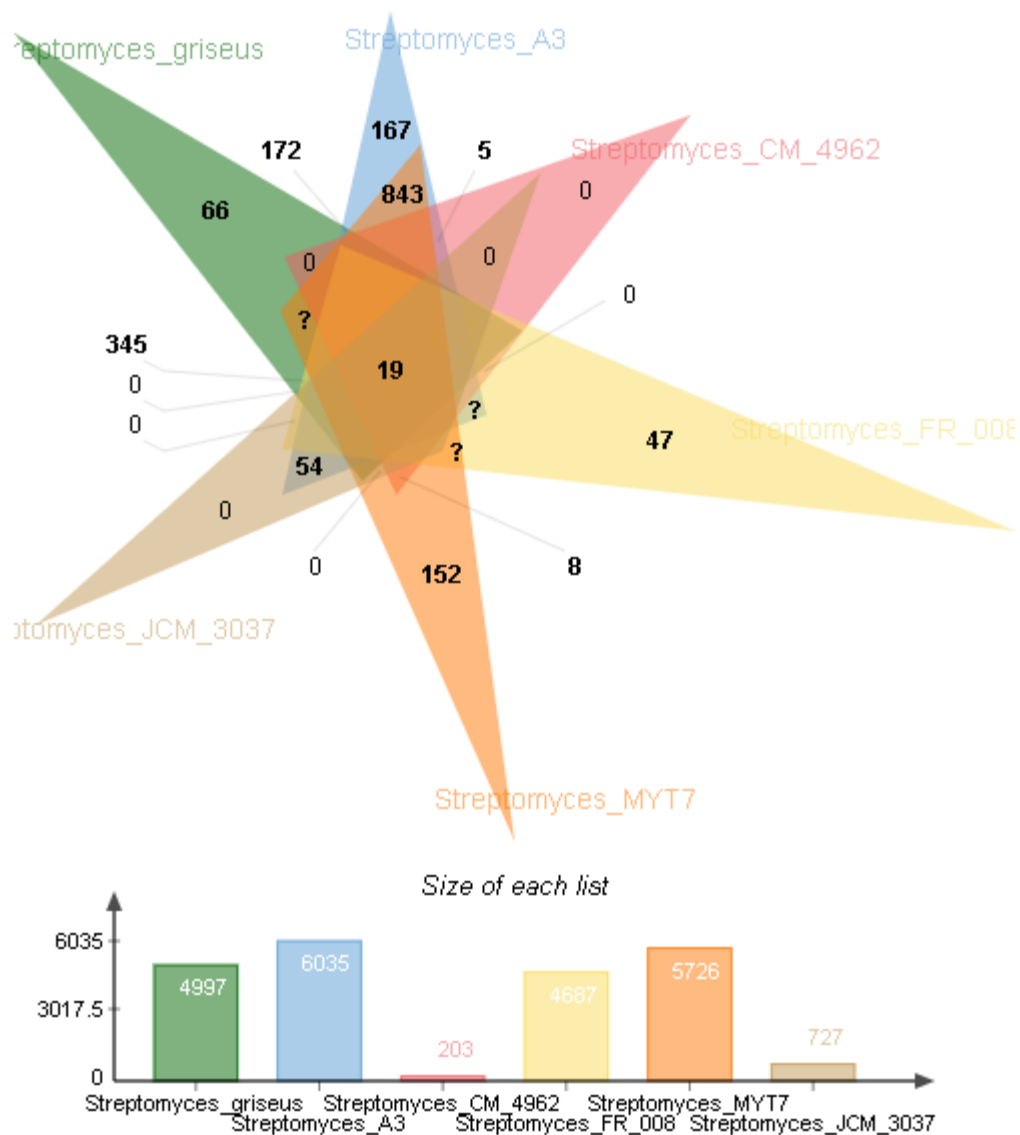


Рис. 29. Діаграма Венна відображає розподіл спільних ортологічних кластерів серед геномів

Як ми бачимо, між цими штамми знайдено 19 спільних ортологічних кластерів. Кластерів, як власиві лише для штаму *Streptomyces ambofaciens* Myt7

виявилось 152 (рис. 30). Серед них ідентифіковано 71 кластер генів, серед яких були ідентифіковані кластери, які відповідають за відомі біологічні процеси, молекулярні функції та за клітинні процеси

Streptomyces ambofaciens Myt7

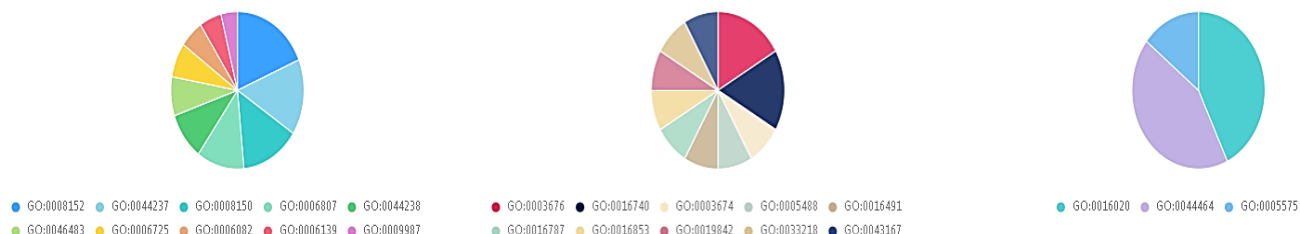


Рис. 30. Ідентифікація біологічних функцій кластеру, присутнього лише в геномі штаму *Streptomyces ambofaciens* Myt7

Перелік ідентифікованих кластерів наведено у таблиці 9.

Таблиця 9

Перелік ідентифікованих кластерів генів штаму *Streptomyces ambofaciens* Myt7 з відомими функціями

Ідентифікований кластер	Кількість ідентифікованих кластерів	Назва процесу, за якій відповідає кластер
Кластери, які відповідають за відомі біологічні процеси		
GO:0008152	42	metabolic process
GO:0044237	35	cellular metabolic process
GO:0008150	32	biological_process
GO:0006807	26	nitrogen compound metabolic process
GO:0044238	22	primary metabolic process
GO:0046483	18	heterocycle metabolic process
GO:0006725	16	cellular aromatic compound metabolic process
GO:0006082	13	organic acid metabolic process
GO:0006139	12	nucleobase-containing compound metabolic process
GO:0009987	9	cellular process
GO:0043170	9	macromolecule metabolic process
GO:0017144	8	drug metabolic process
GO:0006259	7	DNA metabolic process
GO:0032196	7	transposition
GO:0051186	5	cofactor metabolic process
GO:0006629	4	lipid metabolic process
GO:0009116	4	nucleoside metabolic process
GO:0065007	4	biological regulation
GO:0050896	3	response to stimulus

GO:0016070	2	RNA metabolic process
GO:0019748	2	secondary metabolic process
GO:0032502	2	developmental process
GO:0043094	2	cellular metabolic compound salvage
GO:0006066	1	alcohol metabolic process
GO:0006518	1	peptide metabolic process
GO:0006730	1	one-carbon metabolic process
GO:0006766	1	vitamin metabolic process
GO:0016043	1	cellular component organization
GO:0032989	1	cellular component morphogenesis
GO:0043101	1	purine-containing compound salvage
GO:0043603	1	cellular amide metabolic process
GO:0044255	1	cellular lipid metabolic process
GO:0045333	1	cellular respiration
Кластери, які відповідають за молекулярні функції		
GO:0003676	2	nucleic acid binding
GO:0016740	2	transferase activity
GO:0016740	1	transferase activity
GO:0003674	1	molecular_function
GO:0005488	1	binding
GO:0016491	1	oxidoreductase activity
GO:0016787	1	hydrolase activity
GO:0016853	1	isomerase activity
GO:0019842	1	vitamin binding
GO:0033218	1	amide binding
GO:0016740	2	transferase activity
GO:0043167	1	ion binding
Кластери, які відповідають за клітинні процеси		
GO:0016020	3	membrane
GO:0044464	3	cell part
GO:0005575	1	cellular_component

3.8. Біотехнологічна схема дослідження кластерів генів вторинних метаболітів

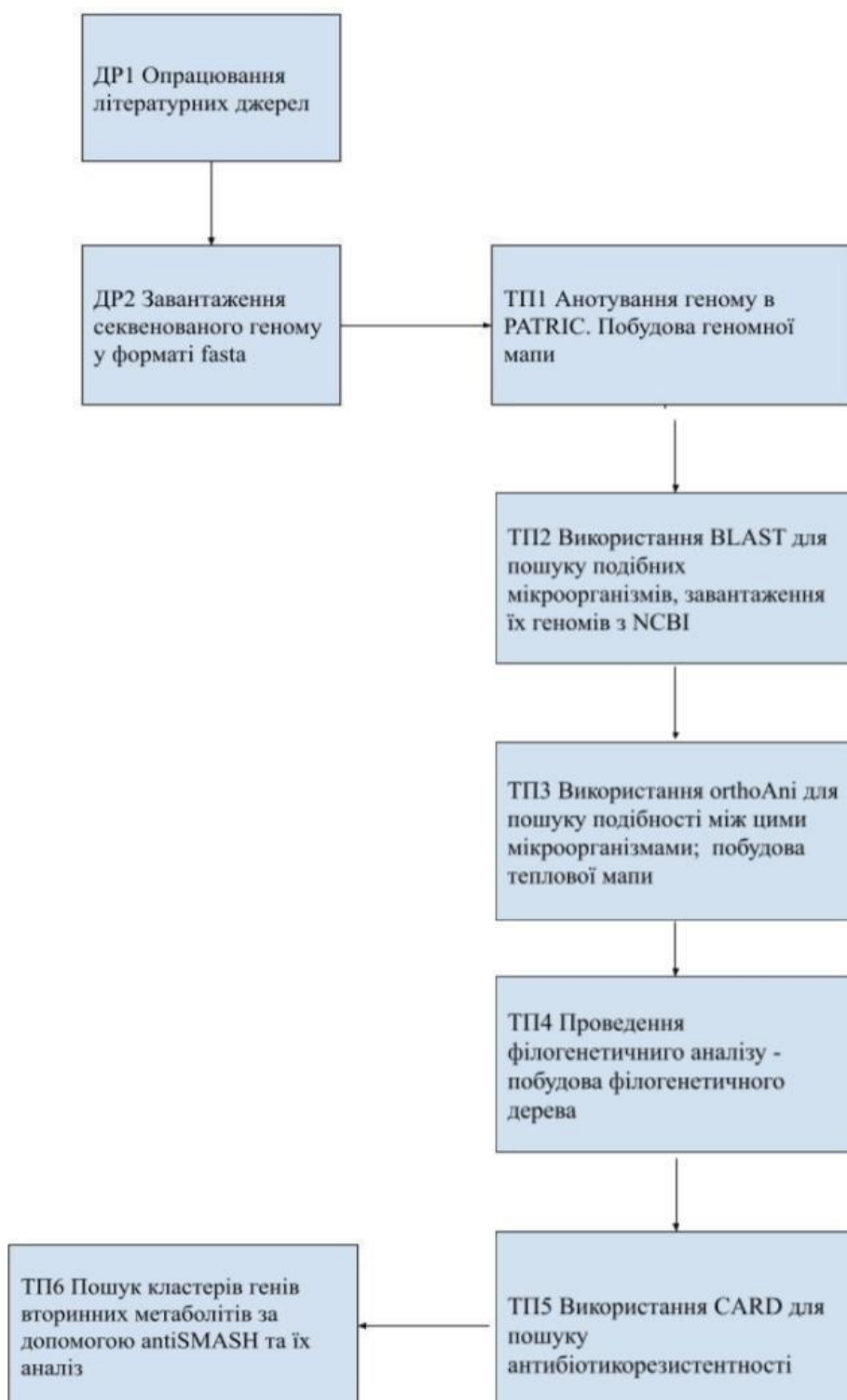


Рис. 31. Біотехнологічна схема дослідження кластерів генів вторинних метаболітів (оригінал)

ДР – допоміжні роботи; ТП - технологічний процес.

ДР1 – опрацювання та аналіз літературних джерел, необхідних для дослідження, зокрема теоретичної інформації щодо характеристики стрептоміцетів, генних кластерів, які пов'язані з продукуванням вторинних метаболітів, активності даних метаболітів та використання програмного забезпечення в роботі.

ДР2 – Завантаження секвенованого геному *Streptomyces ambofaciens* Myt7 у форматі fasta для подальшого вивчення.

ТП1 – Анотування геному в PATRIC, побудова геномної мапи кільцевої хромосоми та візуалізація геному за допомогою PATRIC, побудова геномної мапи із застосуванням серверу CGView Server.

ТП2 – Використання BLAST+ (версія 2.9) для пошуку близькоспоріднених геномів, завантаження їх геномів з NCBI (більшість геномів у вигляді контігів).

ТП3 – Використання orthoANI для визначення ступеня ідентичності між отриманими штамми (через обмеження програми обрали 10 геномів, найбільш споріднених з нашим штамом за результатами попередніх аналізів) та побудова теплової мапи, що демонструє середню ідентичність нуклеотидів за ортологією між штамом *Streptomyces ambofaciens* Myt7 та іншими спорідненими штамми.

ТП4 – Проведення філогенетичного аналізу, тобто побудова філогенетичних дерев, зокрема з використанням серверу TYGS для дерев на підставі ДНК-ДНК гібридизації та програми RAxML на сайті REALPHY для дерев окремих послідовностей генів, знайдених в геномі *Streptomyces ambofaciens* Myt7.

ТП5 – Пошук генів антибіотикорезистентності за допомогою серверу CARD, візуалізація результатів ідентифікації даних генів та ідентифікація бактеріоцинів у сервері BAGEL4.

ТП6 – Пошук кластерів генів вторинних метаболітів нашого штаму за допомогою antiSMASH та їх аналіз, побудова теплової мапи окремих кластерів та пошук ортологічних кластерів генів за допомогою OrthoVenn.

УЗАГАЛЬНЕННЯ

Streptomyces ambofaciens Myt7 представляє високий інтерес для біотехнологічного дослідження, адже є представником морської мікробіоти, яка характеризується унікальними та складними біосинтетичними шляхами продукування вторинних метаболітів за рахунок умов водного середовища.

За результатами аналізу наявності генів антибіотикорезистентності даного штаму на сервері CARD було показано наявність однієї послідовності гену антибіотикорезистентності, знайденою за алгоритмом Strict, що відноситься до макролідних антибіотиків. Через сервер PATRIC було знайдено ще послідовності генів *cmlv* (chloramphenicol resistance), *oleC* (oxytetracycline resistance) та *novA* (efflux ABC транспортер, що був ідентифікований у кластері генів біосинтезу новобіоцину та бере участь у транспорті та стійкості новобіоцину).

При пошуку бактеріоцинів в геномі *Streptomyces ambofaciens* Myt7 з використанням веб-серверу BAGEL4, було ідентифіковано один кластер бактеріоцину – морфогенетичного пептиду SapB, який потрібний для утворення повітряного міцелію актиноміцетів та подібний до лантибіотиків.

В ході біоінформатичного аналізу була виявлена велика кількість генних кластерів, пов'язаних з кодуванням вторинних метаболітів, що мають біологічну активність, зокрема спостерігали наявність NRPS та PKS I, які відповідають за синтез поліоксипептину – гексадепсипептиду, що має активність індукції апоптозу щодо клітин карциноми підшлункової залози людини.

Найбільш часто в геномі зустрічається PKS I, а саме в семи регіонах геному. Наприклад, у регіоні 7 даний кластер кодує синтез віценістатину – полікетидного антибіотику, який має протипухлинну активність щодо карциноми товстої кишки людини. У регіоні 28 PKS I пов'язаний з продукуванням полікетидного антибіотику рифаморфоліну типів А-Е, які мають активність щодо метицилінрезистентного *Staphylococcus aureus*.

Крім того, в геномі було виявлено по одному класу індолу, меланіну, феназіну та ектоїну, що кодують синтез вторинних метаболітів з протигрибковою, протимікробною та противірусною активностями.

Також наявний RiPP-like кластер, що забезпечує продукування рибосомально синтезованого і посттрансляційно модифікованого пептидного продукту, зокрема лантипептиду третього класу інформатипептину.

NRPS-like кластери представлені у регіоні 24 та пов'язані з синтезом стрептотрицину – N-глікозидного антибіотику з активністю щодо грамнегативних бактерій. У регіоні 9 даний кластер забезпечує продукування меридаміцину – неімуносупресивного макроциклічного полікетиду.

Також у геномі *Streptomyces ambofaciens* Myt7 наявна велика кількість кластерів лантипептидів різних класів і терпенів, більшість з яких мають противірусну та антимікробну активність.

Таким чином, дослідження генних кластерів вторинних метаболітів стрептоміцетів Чорного моря, зокрема і *Streptomyces ambofaciens* Myt7, за допомогою методів біоінформатичного аналізу є актуальною і перспективною галуззю для пошуку нових лікарських препаратів.

ВИСНОВКИ

1. За отриманими результатами аналізів з використанням BLAST+, OrthoAni та філогенетичного аналізу, штам *Streptomyces ambofaciens* Myt7 виявився найбільш подібним до штаму *Streptomyces ambofaciens* ATCC 23877.

2. При проведенні біоінформатичного аналізу секвенованого геному штаму *Streptomyces ambofaciens* Myt7, було виявлено 34 регіони геному, пов'язаних з продукуванням вторинних метаболітів та 45 кластерів генів вторинних метаболітів. Найбільшу кількість склали NRPS, PKS та лантипептиди. Було виявлено, що більшість з цих кластерів генів здатні кодувати синтез біологічно активних вторинних метаболітів, зокрема антибіотиків, протипухлинних, антигрибкових та противірусних засобів.

3. Порівнюючі проаналізовані близькоспоріднені геноми з нашим, ми з'ясували, що по кількості та організації BGCs штам *Streptomyces ambofaciens* Myt7 відноситься до одного угруповання зі штамми *Streptomyces ambofaciens* ATCC 23877 та *Streptomyces coelicolor* A3. При цьому, геном штаму *Streptomyces ambofaciens* Myt7 характеризується більшою кількістю кластерів генів вторинних метаболітів наступних типів: PKS I type, PKS III type, NRPS like та RiPP like.

4. Між *Streptomyces ambofaciens* Myt7 та близькоспорідненими штамми, при аналізі з використанням OrthoVenn було знайдено 19 спільних ортологічних кластерів. Кластерів, що властиві лише для нашого штаму виявилось 152, серед них 71 кластерів відповідають за відомі біологічні процеси, молекулярні функції та за клітинні процеси.

СПИСОК ЛІТЕРАТУРИ

1. Пирог Т.П. Загальна мікробіологія / Т.П. Пирог // Систематика прокариот – Київ, 2004. – С. 164 – 168.
2. Сиволоб А.В. Молекулярна біологія / А.В. Сиволоб // Київський університет – Київ, 2008. – С. 109 – 110.
3. Ackerley D.F. Cracking the nonribosomal code. / Ackerley D.F. // Cell Chem Biol – 2016. – Vol. 23. – P. 535 – 537.
4. Agrawal P. RiPPMiner: a bioinformatics resource for deciphering chemical structures of RiPPs based on prediction of cleavage and cross-links/ P. Agrawal, S. Khater, M. Gupta, N. Sain, D. Mohanty // Nucleic Acids Res. – 2017. – V. 45. – P. W80–W88.
5. Aigle B., Sylvie S., Spitteller D., Dickschat J. S., Challis G. L., Leblond P., Pernodet J. // Ind Microbiol Biotechnol – 2013. – Vol. 41 – P. 251 – 263.
6. Aizawa T. Alkyldihydropyrones, new polyketides synthesized by a type III polyketide synthase from *Streptomyces reveromyceticus*. / T. Aizawa, S.Y. Kim, S. Takahashi, M. Koshita, M. Tani, Y. Futamura, H. Osada, N. Funa. // J. Antibiot. – 2014. – Vol. 67. – P. 819–823.
7. Alam K. Complete genome sequencing and in silico genome mining reveal the promising metabolic potential in *Streptomyces* strain CS-7 / Alam K., Hao J., Zhong L., Fan G., Ouyang Q., Islam M. M., Islam S., Sun H., Zhang Y., Li R. // Front Microbiol – 2022. – Vol. 13. – P. 9399.
8. Alcock B.P. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database/ B.P. Alcock, W. Huynh, R. Chalil, K.W. Smith, A.R. Raphenya, M.A. Wlodarski, A. Edalatmand, A. Petkau, S.A. Syed, K.K. Tsang, S.J.C. Baker, M. Dave, M.C. McCarthy, K.M. Mukiri, J.A. Nasir, B. Golbon, H. Imtiaz, X. Jiang, K. Kaur, M. Kwong, Z.C. Liang, K.C. Niu, P. Shan, J.Y.J. Yang, K.L. Gray, G.R. Hoad, B. Jia, T. Bhandu, L.A. Carfrae, M.A. Farha, S. French, R. Gordzevich, K. Rachwalski, M.M. Tu, E. Bordeleau, D. Dooley, E. Griffiths, H.L. Zubyk, E.D. Brown, F. Maguire, R.G. Beiko, W.W.L. Hsiao, F.S.L., Brinkman G. Van

- Domselaar, McArthur A. G. // *Nucleic Acids Res.* – 2023. – V. 6, No. 51. – P. D690-D699.
9. Altenhoff A.M. The OMA orthology database in 2018: retrieving evolutionary relationships among all domains of life through richer web and programmatic interfaces/ A.M. Altenhoff, N.M. Glover, C.-M. Train, K. Kaleb, V.A. Warwick, D. Dylus, T.M. de Farias, K. Zile, C. Stevenson, J. Long // *Nucleic Acids Res.* – 2018. – V. 46. – P. D477–D485.
 10. Auch A.F. Standard operating procedure for calculating genome-to-genome distances based on high-scoring segment pairs/ A.F. Auch, H.P. Klenk, M. Göker // *Stand Genomic Sci.* – 2010. – V. 2. – P. 142-148.
 11. Barka E. Taxonomy, physiology, and natural products of Actinobacteria. / Barka E.; Vatsa P.; Sanchez L.; Gaveau-Vaillant N.; Jacquard C.; Klenk H; Clément C.; Ouhdouch Y.; Van Wezel G. // *Microbiol. Mol Biol. Rev* – 2016. – P. 1 – 43.
 12. Bédard F., Biron E. Recent progress in the chemical synthesis of class II and S-glycosylated bacteriocins/ Bédard F., Biron E. // *Front Microbiol.* – 2018. – V. 9. – P. 1–4.
 13. Berdy J. Thoughts and facts about antibiotics: Where we are now and where we are heading. / Berdy J. // *Antibiot* – 2012. – Vol. 65 – P. 385 – 395.
 14. Blin K. AntiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification / Blin K., Wolf T., Chevrette M.G., Lu X., Schwalen C.J., Kautsar S.A., Suarez Duran H.G., de Los Santos E.L.C., Kim H.U., Nave M. // *Nucleic Acids Res.* – 2017. – Vol. 45. – P. 36 – 41.
 15. Blin K. antiSMASH 6.0: improving cluster detection and comparison capabilities / Blin K., Shaw S., Kloosterman A. M., Charlop-Powers Z., Wezel G. P., Medema M. H., Weber T. // *Nucleic Acids Res* – 2021. – Vol. 49. – P. 29 – 35.
 16. Capstick D. S. SapB and the chaplins: connections between morphogenetic proteins in *Streptomyces coelicolor* /D. S. Capstick, J.M. Willey, M. J. Buttner, M. A. Elliot// *Molecular Microbiology.* – 2007. – V. 64, No. 3. – P. 602–613
 17. Chen C.Y. Detection of Antimicrobial Resistance Using Proteomics and the Comprehensive Antibiotic Resistance Database: A Case Study/ C.Y. Chen, C.G.

- Clark, S. Langner, D.A. Boyd, A. Bharat, S.J. McCorrister, A.G. McArthur, M.R. Graham, G.R. Westmacott, G. Van Domselaar // *Proteomics Clin Appl.* – 2020. – V. 14, No. 4. – Article: e1800182
18. Chew W. K. Significant reduction of brain cysts caused by *Toxoplasma gondii* after treatment with spiramycin coadministered with metronidazole in a mouse model of chronic toxoplasmosis / Chew W. K., Segarra I., Ambu S., Mak J. W. // *Antimicrob Agents Chemother* – 2012. – Vol. 56 – P. 1762 – 1768.
 19. Chun J. Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes/ J. Chun, A. Oren, A. Ventosa, H. Christensen, D.R. Arahal, M.S. Costa, A.P. Rooney, H. Yi, X.W. Xu, S. De Meyer // *Int. J. Syst. Evol. Microbiol.* – 2018. – V. 68. – P. 461–466.
 20. Cosar C. Trypanocide action of an antibiotic produced by a *Streptomyces*. Cosar C., Ninet L., Pinnert-Sindico S., Preud'Homme J. // *C R Hebd Seances Acad Sci* – 1952. – Vol. 234 – P. 1498 – 1499.
 21. Cotter P.D. Bacteriocins: developing innate immunity for food./ P.D. Cotter, C. Hill, R.P. Ross // *Nat Rev Microbiol.* – 2005. – V. 3. – P. 777–788.
 22. Dayu Y. Type III polyketide synthases in natural product biosynthesis / Dayu Y., Fuchao X., Jia Z., Jixun Z. // *IUBMB Life* – 2012. – Vol. 64 – P. 285 – 295.
 23. de Jong W. SapB and the rodlinins are required for development of *Streptomyces coelicolor* in high osmolarity media /W. de Jong, E. Vijgenboom, L. Dijkhuizen, H. A. B. Wösten, D. Claessen// *FEMS Microbiol Lett.* – 2012. – V. 329, No. 2. – P. 154-159.
 24. Delcher A.L. Identifying bacterial genes and endosymbiont DNA with Glimmer/ A.L. Delcher, K.A. Bratke, E.C. Powers, S.L. Salzberg // *Bioinformatics.* – 2007. – V. 23. –P. 673–679.
 25. Du L. PKS and NRPS release mechanisms / Du L., Lou L. // *Nat Prod Rep.* – 2010. – Vol. 27. – P. 255 – 278.
 26. Emms D.M., Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy/ D.M. Emms, S. Kelly // *Genome Biol.*– 2015. – V. 16. – Article:157.

27. Fang G. Getting started in gene orthology and functional analysis/ G. Fang, N. Bhardwaj, R. Robilotto, M.B. Gerstein // PLoS Comput. Biol. – 2010. – V. 6.– Article: e1000703
28. Field D. Developing bacteriocins of lactic acid bacteria into next generation biopreservatives/ D. Field, R.P. Ross, C. Hill // Curr Opin Food Sci. – 2018. – V. 20. – P. 1–6
29. Fischbach M. A. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms / Fischbach MA, Walsh CT // Chem Rev. – 2006. – Vol. 106 – P. 3468 – 3496.
30. Fouts D.E. PanOCT: automated clustering of orthologs using conserved gene neighborhood for pan-genomic analysis of bacterial strains and closely related species/ D.E. Fouts, L. Brinkac, E. Beck, J. Inman, G. Sutton // Nucleic Acids Res. – 2012. – V. 40. – Article:e172
31. Funabashi M. Phenolic lipids synthesized by type III polyketide synthase confer penicillin resistance on *Streptomyces griseus*. / M. Funabashi, N. Funa, S. J. Horinouchi. //Biol. Chem. – 2008.–Vol. 283. –P. 13983– 13991.
32. Funa N. Biosynthesis of hexahydroxyperylenequinone melanin via oxidative aryl coupling by cytochrome P-450 in *Streptomyces griseus*. / Funa N., Funabashi M., Ohnishi Y., Horinouchi S. // J. Bacteriol. – 2005. – Vol. 187 – P. 8149 – 8155.
33. Gabrielsen C., Brede D.A., Nes I.F. Circular bacteriocins: biosynthesis and mode of action/ C. Gabrielsen, D.A. Brede, I.F. Nes // Appl Environ Microbiol. – 2014. – V. 80. – P. 6854–6862.
34. Gong W. Solution Structure of LCI, a Novel Antimicrobial Peptide from *Bacillus subtilis*/ W. Gong, J. Wang, Z. Chen, B. Xia, G. Lu// Biochemistry. – 2011. – V. 50, No. 18. – P. 3621–3627
35. Jia B. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database/ B. Jia, A.R. Raphenya, B. Alcock, N. Waglechner, P. Guo, K.K. Tsang, B.A. Lago, B.M. Dave, S. Pereira, A.N. Sharma //Nucleic Acids Res. – 2017. – V. 45. – P. D566–D573.

36. Jia B. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database/B. Jia., A.R. Raphenya, B. Alcock, N. Waglechner// *Nucleic Acids Research*. – 2017. – V. 45. – P. D566–D573.
37. Juguet M. An iterative nonribosomal peptide synthetase assembles the pyrrole-amide antibiotic congocidine in *Streptomyces ambofaciens* // Juguet M., Lautru S., Francou F. X., Nezbedová S., Leblond P., Gondry M., Pernodet J. L. // *Chem Biol*. – 2009. – Vol. 16 – P. 421 – 431.
38. Kim M. Transcriptomics-based strain optimization tool for designing secondary metabolite overproducing strains of *Streptomyces coelicolor* / Kim M., Yi J. S., Lakshmanan M., Lee D. Y., Kim B. G. // *Biotechnol. Bioeng*. – 2016. – Vol. 113 – P. 651 – 660.
39. Kingsford C.L. Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake/ C.L. Kingsford, K. Ayanbule, S.L. Salzberg // *Genome Biol*. – 2007. – V. 8. – Article: R22.
40. Kocot K.M. PhyloTreePruner: a phylogenetic tree-based approach for selection of orthologous sequences for phylogenomics/ K.M. Kocot, M.R. Citarella, L.L. Moroz, K.M. Halanych // *Evol. Bioinform. Online*. – 2013. – V. 9. – P. 429–435
41. Kristensen D.M. Computational methods for Gene Orthology inference / D.M. Kristensen, Y.I. Wolf, A.R. Mushegian, E.V. Koonin// *Brief. Bioinform*. – 2011. – V. 12. – P. 379–391.
42. Lee I. OrthoANI: an improved algorithm and software for calculating average nucleotide identity/ I. Lee, Y.O. Kim, S.C. Park, J. Chun // *Int J Syst Evol Microbiol*. – 2016. – V. 66. – P. 1100 –1103.
43. Louca S. A census-based estimate of earth's bacterial and archaeal diversity/ S. Louca, F. Mazel, M. Doebeli, L.W. Parfrey // *PLoS Biol*. – 2019. – V. 17. – Article:e3000106.
44. Ludwig W. Road map of the phylum Actinobacteria/ W. Ludwig, J. Euzéby, P. Schumann, H. Busse, M.E. Trujillo, P. Kämpfer, W.B. Whitman // *Bergey's Manual of Systematics of Archaea and Bacteria*. – 2012. – Vol. 5. – P. 1 – 28.

45. Madden D.E. Taking the next-gen step: Comprehensive antimicrobial resistance detection from *Burkholderia pseudomallei*./ D.E. Madden, J.R. Webb, E.J. Steinig, B.J. Currie, E.P. Price, D.S. Sarovich // EBioMedicine. – 2021. – V. 63. – Article: 103152.
46. McArthur A.G. The comprehensive antibiotic resistance database/ A.G. McArthur, N. Waglechner, F. Nizam, A. Yan, M.A. Azad, A.J. Baylay, K. Bhullar, M.J. Canova, G. De Pascale, L. Ejim // Antimicrob. Agents Chemother. – 2013. – V. 57. – P. 3348–3357.
47. Medema M.H. AntiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences Medema M.H., Blin K., Cimermancic P., de Jager V., Zakrzewski P., Fischbach M.A., Weber T., Takano E., Breitling R. // Nucleic Acids Res. – 2011. – Vol. 39. – P. 339 – 346.
- a. Medema M.H. Minimum information about a biosynthetic gene cluster/ M.H. Medema, R. Kottmann, P. Yilmaz, M. Cummings, J.B. Biggins, K. Blin, I. de Bruijn, Y.H. Chooi, J. Claesen, R.C. Coates // Nat. Chem. Biol. – 2015. – V. 11. – P. 625–631.
48. Meier-Kolthoff J.P. Genome sequence-based species delimitation with confidence intervals and improved distance functions/ J.P. Meier-Kolthoff, A.F. Auch, H.-P. Klenk, M. Göker // BMC Bioinformatics. – 2013. – V. 14. – Article:60
49. Meier-Kolthoff J. P. TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes /J. P. Meier-Kolthoff, J. S. Carbasse, R. L. Peinado-Olarte, M. Göker// Nucleic Acids Research. – 2022. – V. 50, Issue D1. – P. D801–D807.
50. Miyanaga A. Direct transfer of starter substrates from type I fatty acid synthase to type III polyketide synthases in phenolic lipid synthesis. / A. Miyanaga, N. Funa, T. Awakawa, S. Horinouchi //Proc. Natl. Acad. Sci. USA. – 2008. – Vol. 105. – P. 871– 876.

51. Mullowney M.W. Natural products from thioester reductase containing biosynthetic pathways. / Mullowney M.W., McClure R.A., Robey M.T., Kelleher N.L., Thomson R.J. // Nat. Prod. Rep. – 2018. – Vol. 35 – P. 847 – 878.
52. Page A.J. Roary: rapid large-scale prokaryote pan genome analysis/ A.J. Page, C.A. Cummins, M. Hunt, V.K. Wong, S. Reuter, M.T.G. Holden, M. Fookes, D. Falush, J.A. Keane, J. Parkhill.// Bioinformatics. –2015. – V. 31. – P. 3691–3693
53. Pang X. Functional angucycline-like antibiotic gene cluster in the terminal inverted repeats of the *Streptomyces ambofaciens* linear chromosome / Pang X., Aigle B., Girardet J., Mangenot S., Pernodet J., B. Decaris B., Leblond P. // Antimicrob. Chemother. – 2004. – Vol. 48 – P. 575 – 588.
54. Piel J. Biosynthesis of polyketides by trans-AT polyketide synthases. / Piel J. // Nat Prod Rep. – 2010. – P. 996 –1047.
55. Poulet P. P. Concentrations and in vivo antibacterial activity of spiramycin and metronidazole in patients with periodontitis treated with high-dose metronidazole and the spiramycin/metronidazole combination / Poulet P. P., Duffaut D., Barthet P., Brumpt I. // Antimicrob. Chemother. – 2005. – Vol. 55 – P. 347 – 351.
56. Rashid R. Focal targeting of the bacterial envelope by antimicrobial peptides / R. Rashid, M. Veleba, K. A. Kline // Front. Cell. Dev. Biol. – 2016. – V. 4. – P. 55.
57. Reimer, J. M. Structures of a dimodular nonribosomal peptide synthetase reveal conformational flexibility. / Reimer, J. M. // Science – 2019. – Vol. 366 – P. 4388.
58. Robbins T. Structure and mechanism of assembly line polyketide synthases. / Robbins T., Liu Y.C., Cane D.E., Khosla C. // Curr Opin Struct Biol. – 2016. – P. 8 –10.
59. Rodríguez A.M. *Streptomyces antibioticus* contains at least three oleandomycin-resistance determinants, one of which shows similarity with proteins of the ABC-transporter superfamily/ A.M. Rodríguez, C. Olano, C. Vilches, C. Méndez, J.A. Salas //Mol Microbiol. – 1993. – V. 8, No. 3. – P. 571-82.
60. Saikia K. Aromatic-rich C-terminal region of LCI is a potent antimicrobial peptide in itself/ K. Saikia, V. K. Belwal, D. Datta, N. Chaudhary //Biochemical

- and Biophysical Research Communications. – 2019. – V.519, Issue 2. – P. 372-377.
61. Schleifer K. Classification of Bacteria and Archaea: past, present and future/ Schleifer K. // *Syst Appl Microbiol.* – 2009. – V. 32, No. 8. – P. 533-542.
 62. Schmutz E. Resistance genes of aminocoumarin producers: two type II topoisomerase genes confer resistance against coumermycin A1 and clorobiocin/ E. Schmutz, A. Mühlenweg, S.M. Li, L. Heide // *Antimicrob Agents Chemother.* – 2003. – V. 47, No. 3. – P. 869-877
 63. Schreiber F., Patricio M., Muffato M., Pignatelli M., Bateman A.. TreeFam v9: a new website, more species and orthology-on-the-fly/ F. Schreiber, M. Patricio, M. Muffato, M. Pignatelli, A. Bateman.//*Nucleic Acids Res.* – 2014. – V. 42. – P. D922–D925
 64. Shimizu Y. Type III polyketide synthases: functional classification and phylogenomics. / Shimizu Y., Ogata H, Goto S. // *ChemBioChem.* – 2017. – P. 50 –65.
 65. Stackebrandt E. Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology/ E. Stackebrandt, B.M. Goebel// *Int J Syst Bacteriol.* – 1994. – V. 44, No. 4. – P. 846-849.
 66. Stumpp T. Cloning of the netropsin resistance genes from *Streptomyces flavopersicus* NRRL 2820 / Stumpp T., Himbert S., Altenbuchner J. // *Basic Microbiol* – 2005. – Vol. 45 – P. 355 – 362.
 67. Sumi C. D. Antimicrobial peptides of the genus *Bacillus*: a new era for antibiotics / Sumi C. D., Yang B. W., Yeo I. C., Hahm Y. T. // *Can J Microbiol* – 2015. – Vol. 61 – P. 93 – 103.
 68. Tang X. A two-step sulfation in antibiotic biosynthesis requires a type III polyketide synthase. / X. Tang, K. Eitel, L. Kaysser, A. Kulik, S. Grond, B. Gust // *Nat. Chem. Biol.* – 2013. –Vol 9. – P. 610–615.
 69. Tarry X-Ray crystallography and electron microscopy of cross- and multi-module nonribosomal peptide synthetase proteins reveal a flexible architecture /

- Tarry, M.J., Haque, A.S., Bui, K.H, Schmeing, T.M. // *Structure* – 2017. – Vol. 25 – P. 783 – 793.
70. Tindall B.J. Notes on the characterization of prokaryote strains for taxonomic purposes/ B.J. Tindall, R. Rosselló-Móra, H.-J. Busse, W. Ludwig, P. Kämpfer // *Int. J. Syst. Evol. Microbiol.* – 2010. – V. 60. – P. 249–266.
71. Tseng C. C. Role of the active site cysteine of DpgA, a bacterial type III polyketide synthase. /A. Miyanaga, N. Funai, T. Awakawa, S. Horinouchi // *Biochemistry.* – 2004. – Vol. 43. – P. 970– 980.
72. van Heel A.J. BAGEL4: a user-friendly web server to thoroughly mine RiPPs and bacteriocins /A.J. van Heel, A. de Jong, C. Song, J. H. Viel, J. Kok, O. P. Kuipers // *Nucleic Acids Research.*– 2018. – V. 46, Issue W1. – P. W278–W281.
73. Vingadassalon A. Transcriptional Regulation of Congocidine (Netropsin) Biosynthesis and Resistance / Vingadassalon A., Lorieux F., Juguet M., Noël A., Santos L. D., Fernandez L. M., Pernodet J., Bury-Moné S., Lautru S. // *Appl Environ Microbiol* – 2021. – Vol. 87 – P. 138021.
74. Walsh C.T. Insights into the chemical logic and enzymatic machinery of NRPS assembly lines. / Walsh C.T. // *Nat. Prod. Rep.* – 2016. – Vol. 33 – P. 127 – 135.
75. Wenzel S.C. A type I/type III polyketide synthase hybrid biosynthetic pathway for the structurally unique ansa compound kendomycin. / Wenzel S.C., Bode H.B., Kochems I., Müller R. // *ChemBioChem* – 2008. – Vol. 9 – P. 2711 – 2721.
76. Wu L. Bacilysin overproduction in *Bacillus amyloliquefaciens* FZB42 markerless derivative strains FZBREP and FZBSPA enhances antibacterial activity/ L. Wu, H. Wu, L. Chen, L. Lin, R. Borriss, X. Gao // *Appl Microbiol Biotechnol.* – 2015. – V. 99, No. 10. – P. 4255–4263.
77. Xu L. OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species/ L. Xu, Z. Dong, L. Fang, Y. Luo, Z. Wei, H. Guo, G. Zhang, Y.Q. Gu, D. Coleman-Derr, Q. Xia, Y. Wang // *Nucleic Acids Res.* – 2019. – V. 2, No. 47. – P. W52-W58

78. Xu W. Structural analysis of protein-protein interactions in type I polyketide synthases. / Xu W, Qiao K, Tang Y // Crit Rev Biochem Mol Biol. – 2013. – P. 98 – 122.
79. Zdobnov E.M. OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs/ E.M. Zdobnov, F. Tegenfeldt, D. Kuznetsov, R.M. Waterhouse, F.A. Simão, P. Ioannidis, M. Seppey, A. Loetscher, E.V. Kriventseva // Nucleic Acids Res. – 2017. – V. 45. – P. D744–D749.
80. Zha W. Characterization of the substrate specificity of PhlD, a type III polyketide synthase from *Pseudomonas fluorescens*. / W. Zha, S. B. Rubin-Pitel, H. Zhao // J. Biol. Chem. – 2008. – Vol. 281. – P. 32036 – 32047.