

ЮЩУК О.С.¹, ЖУКРОВСЬКА К.А.¹, ФЕДОРЕНКО В.О.^{1✉}

Львівський національний університет імені Івана Франка,
Україна, 79005, м. Львів, вул. Грушевського, 4

✉ viktor.fedorenko@lnu.edu.ua, +38(032) 239-44-07, +38(032) 239-44-75

ПОШИРЕННЯ КЛАСТЕРІВ ГЕНІВ БІОСИНТЕЗУ КОМПЛЕСТАТИНУ ТА ПОДІБНИХ СПОЛУК У ПРЕДСТАВНИКІВ РОДУ *STREPTOMYCES*

Мета. У цій роботі проаналізовано 2664 геноми представників роду *Streptomyces* (послідовності яких доступні в базі даних GenBank) з метою пошуку кластерів біосинтетичних генів (КБГ), подібних до КБГ комплестатину. На основі послідовностей знайдених КБГ було реконструйовано біосинтетичні шляхи, що дозволило оцінити структурну різноманітність комплестатин-подібних сполук, які походять із природних джерел. **Методи.** В роботі застосовано набір загальноновживаних програм для пошуку КБГ та філогенетичної реконструкції. **Результати.** Комплестатин-подібні КБГ було знайдено в 53 геномах, хоча лише 33 локалізувалися в межах одного контігу – інші КБГ були неповні або фрагментовані. Реконструкція мультилокусної філогенії 33 знайдених повних КБГ та КБГ комплестатину дозволила виокремити п'ять філогенетичних підгруп. Представники кожної підгрупи мали характерну генетичну організацію КБГ. Реконструкція потенційних біосинтетичних шляхів, кодованих у знайдених КБГ, показала, що ці КБГ можуть кодувати біосинтез деяких нових похідних комплестатину, а саме норкомплестатину, *N*-малоніл-норкомплестатину та *N*-ацетил-норкомплестатину. **Висновки.** Комплестатин-подібні КБГ є розповсюдженими серед стрептоміцетів та можуть кодувати біосинтез нових похідних комплестатину, що заслуговує подальшого експериментального вивчення.

Ключові слова: комплестатин, кластери генів біосинтезу, стрептоміцети, майнінг геномів.

Глікопептидні антибіотики (ГПА) відомі як потужні інгібітори синтезу клітинної стінки Грам-позитивних бактерій, що зв'язують ліпід II [1]. Сьогодні в клініці використовуються два природні ГПА, а також три напівсинтетичні сполуки. До перших належить ванкомицин і тейкопланін, що десятками років ефективно використовуються в клініці для боротьби із мультирезистентними Грам-позитивними патогенами [1].

ГПА не вичерпуються сполуками, що використовуються в медицині як протибактерійні препарати. Описано десятки різноманітних ГПА, які мали обмежене медичне застосування, або взагалі не використовувалися [1]. Усі ГПА класифікують на п'ять типів залежно від особливостей їхньої структури [1]. До типу I належать глікозильовані гептапептиди, що містять аліфатичні амінокислоти у 1 та 3 положеннях аглікону і три поперечні зшивки між ароматичними кільцями амінокислот у положеннях 2 і 4, 4 і 6, 5 і 7. Представники типу II – це глікозильовані сполуки, в яких, на відміну від ГПА типу I, всі сім амінокислот їхніх агліконів ароматичні. В агліконах ГПА типу III всі сім амінокислот ароматичні і характеризуються одною додатковою поперечною зшивкою між ароматичними кільцями амінокислот у 1 та 3 положеннях. До цього типу належать як глікозильовані, так і неглікозильовані сполуки. За структурою аглікону ГПА типу IV аналогічні до типу III, однак містять аліфатичний бічний ланцюг. ГПА типу V не глікозильовані, можуть бути як гепта- так і нона-пептидами, містять різні поперечні зшивки. У строгому сенсі, ГПА типу V не є глікопептидами. Більше того, ці сполуки не здатні зв'язувати ліпід II, а крім антимікробних, для них характерні й інші біологічні активності, важливі з фармакологічної точки зору [1].

Одним із цікавих ГПА V-го типу є комплестатин, вперше виділений з *Streptomyces lavendulae* SANK 60477 [2]. Комплестатину властива слабка антимікробна активність. Донедавна вважалося, що він інгібує біосинтез жирних кислот Грам-позитивних бактерій, але більш детальний аналіз показав, що молекулярними мішенями комплестатину є аутолізини – ферменти ремоделювання та утилізації муреїну бактерій [3]. Комплестатин здатний ефективно пригнічувати формування біоплівки Грам-негативними бактеріями через пригнічення нітритредуктази [4]. Однак, найважливішим є те, що комплестатин виявився потужним інгібітором

© ЮЩУК О.С., ЖУКРОВСЬКА К.А., ФЕДОРЕНКО В.О.

гемолізу еритроцитів, сенсibiliзованих під впливом системи комплементу [2]. Комплестатин також здатний *in vitro* інгібувати розвиток вірусу імунодефіциту людини HIV-1 [5]. Варто також відзначити, що комплестатин є потужним протектором нейронів від ексайтотоксичності, що спричинена *N*-метил-D-аспаратом (NMDA) та α -аміно-3-гідрокси-5-метил-4-ізоксазолпропіоною кислотою (AMPA). Така нейропротекторна дія комплестатину є наслідком блокування транспорту іонів Ca^{2+} після активації NMDA- та AMPA-рецепторів. Більше того, комплестатин демонстрував здатність захищати нейрони кори головного мозку від кисневого та глюкозного голодування [6]. Для комплестатину також описані певні протиракові активності [7]. Кластер біосинтетичних генів (КБГ) комплестатину з *S. lavendulae* SANK 60477 (*com*-кластер) секвеновано одним із перших серед генів ГПА [8]. Згодом показано наявність КБГ комплестатину в *S. chartreusis* AN1542 [9], *Streptomyces* sp. WAC01325 і деяких інших стрептоміцетів [10]. Можна припустити, що КБГ комплестатину є доволі поширеними серед представників роду *Streptomyces*. Однак, дотепер не проводився порівняльний геномний аналіз *com*-подібних КБГ, їх поширення та особливостей організації. Тому, метою нашої роботи був аналіз нуклеотидних послідовностей 2664 геномів стрептоміцетів, доступних сьогодні в базі даних GenBank, на наявність КБГ, подібних до *com*-кластера. Порівняння послідовностей КБГ дасть змогу оцінити різноманітність ГПА V-го типу, подібних до комплестатину, та спростить пошук нових антибіотиків.

Матеріали і методи

Скринінг геномів стрептоміцетів у пошуках КБГ, подібних до комплестатину. Нуклеотидні послідовності 2664 геномів *Streptomyces* spp. було отримано із репозиторію GenBank у вигляді файлів *.gb. Перелік усіх цих послідовностей можна знайти онлайн у Електронному Додатку 1 (ЕД1). Для пошуку *com*-подібних КБГ у відібраних нуклеотидних послідовностях було використано наступний алгоритм дій: 1) за допомогою програми MultiGeneBlast, <http://multigeneblast.sourceforge.net> [11] створено offline базу даних, що включала нуклеотидні послідовності з ЕД1; 2) конкатеновані послідовності всіх генів *com* (AF386507, крім генів нерибосомних пептидсинтез – НРПС) використано як «запит» для пошуку проти створеної бази

даних в MultiGeneBlast; 3) нуклеотидні послідовності, в яких було виявлено ко-локалізовані гомологи *com*-генів, проаналізовано на наявність КБГ антибіотиків за допомогою antiSMASH 6.0 [12]; 4) отримані послідовності *com*-подібних кластерів перевірено вручну і за потреби переанотовано.

Філогенетична реконструкція. Пакет програм MEGA 10 [13] використано для створення множинних вирівнювань конкатенованих амінокислотних послідовностей білків, закодованих в 33 із знайдених КБГ, що є ортологами семи білків закодованих в *com*, а саме: AAK81822, AAK81823, AAK81829, AAK818230, AAK81831 та AAK81832. Ці послідовності можна знайти онлайн в Електронному Додатку 2 (ЕД2). MEGA 10 також використали для визначення оптимальної моделі філогенії, застосованої для реконструкції філогенетичного дерева методом максимальної правдоподібності. Такою виявилася модель Jones-Taylor-Thornton (JTT) [14] з оцінкою γ -розподілу варіацій частот між сайтами (5 дискретних γ -категорій, G5) та емпіричними частотами амінокислот (F). Достовірність топології філогенетичного дерева, побудованого за цією моделлю, була оцінена за допомогою бутстреп-тесту (1000 повторів).

Для не згаданих вище рутинних маніпуляцій та аналізів амінокислотних та нуклеотидних послідовностей використовували GENEIOUS 4.8.5 [15].

Результати та обговорення

Розповсюдження com-подібних КБГ в геномах стрептоміцетів. Ми проаналізували нуклеотидні послідовності 2664 геномів представників роду *Streptomyces*. Серед них були як повністю зібрані геноми, так і чернетки різної якості. КБГ, подібні до *com*, знайдено в 53 геномах. Так як було проаналізовано нуклеотидні послідовності різної якості, лише у 33 випадках ми отримали повні послідовності КБГ (див. Електронний Додаток 3, ЕД3) – розміщені у межах одного контігу. У решті 20 випадків КБГ були фрагментовані, або містили помилки секвенування (ЕД3). У цілому, можна вважати, що *com*-подібні КБГ є доволі поширеними у геномах стрептоміцетів. Для дальшого аналізу ми використали тільки ті 33 КБГ, які знайдені в межах одного контігу.

Особливості організації знайдених com-подібних КБГ. Для того, щоб оцінити різномані-

тність організації 34 *com*-подібних КБГ (включно із *com*), ми спершу вирішили реконструювати мульти-локусну філогенію цих КБГ, щоб виявити кореляцію такого філогенетичного фреймворку з особливостями їх організації. З'ясовано, що всі 34 КБГ діляться на 5 добре відділених філогенетичних підгруп (рис. 1).

Аналіз організації КБГ кожної з підгруп показав, що початкове припущення було правильним, і кожна з філогенетичних підгруп включає в себе кластери з характерною організацією. Далі охарактеризовано КБГ підгруп у порядку від базальної підгрупи V до кронувої підгрупи I (рис. 1).

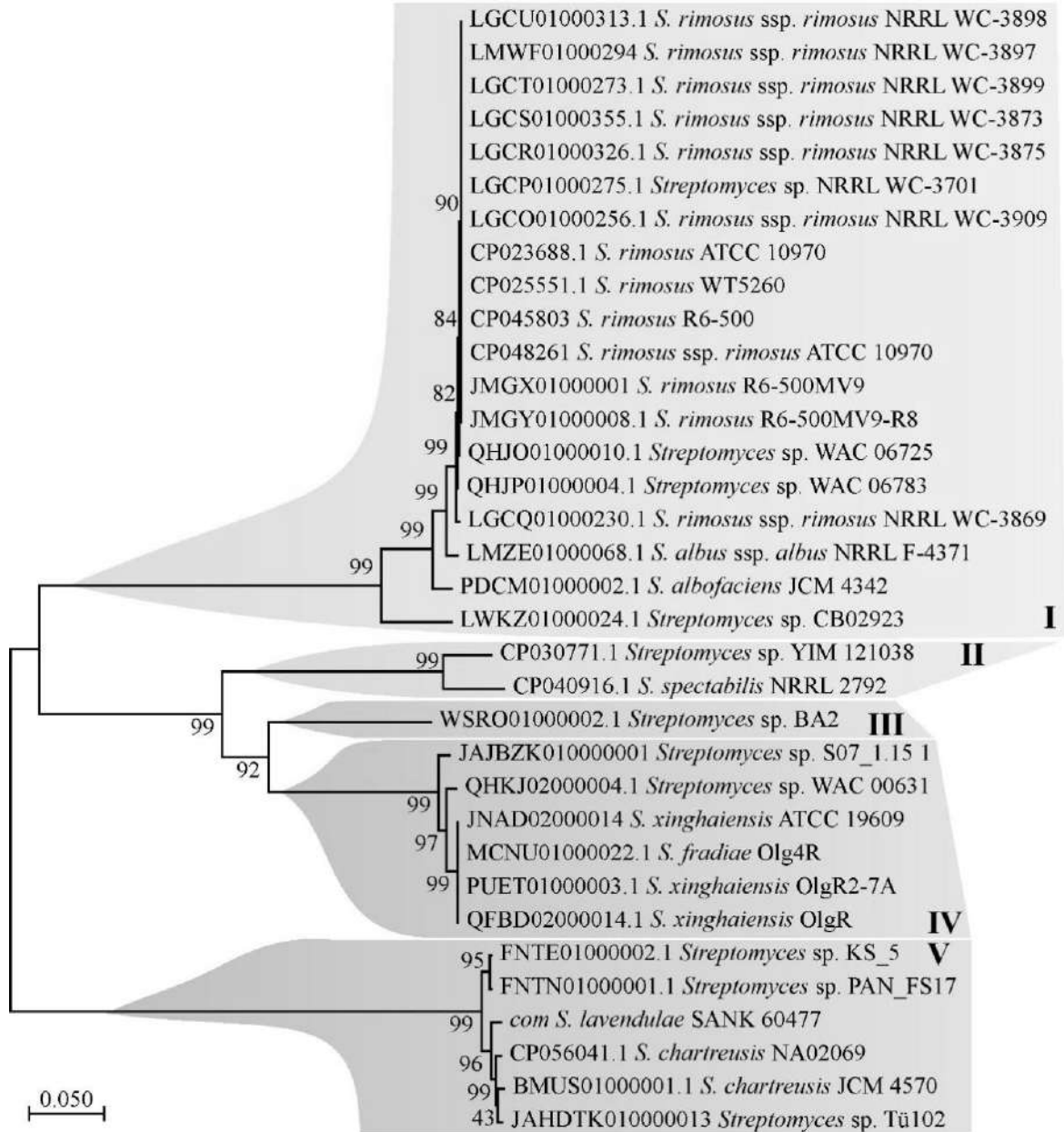


Рис. 1. Філогенетичне дерево конкатенованих амінокислотних послідовностей семи білків (див. розділ «Методи»), що закодовані в 33 *com*-подібних КБГ ГПА, а також в самому *com*-кластері. На дереві виділено 5 клад із високою бутстреп-підтримкою, що відповідають п'яти філогенетичним підгрупам КБГ. Алгоритм побудови цього дерева описано в розділ «Методи». Планка масштабу показує кількість амінокислотних замін на один сайт.

Підгрупа V. До підгрупи V входить *com*-кластер. Усі кластери цієї підгрупи організовані подібно до нього (рис. 2а). Виявилось, що КБГ підгрупи V містять у своєму складі 6 генів, раніше не описаних в кластері *com*. Це два гени

двох-субодиничного АВС-транспортера, два гени двох-компонентної регуляторної системи, а також два гени, що кодують білки, які не задіяні у біосинтезі ГПА.

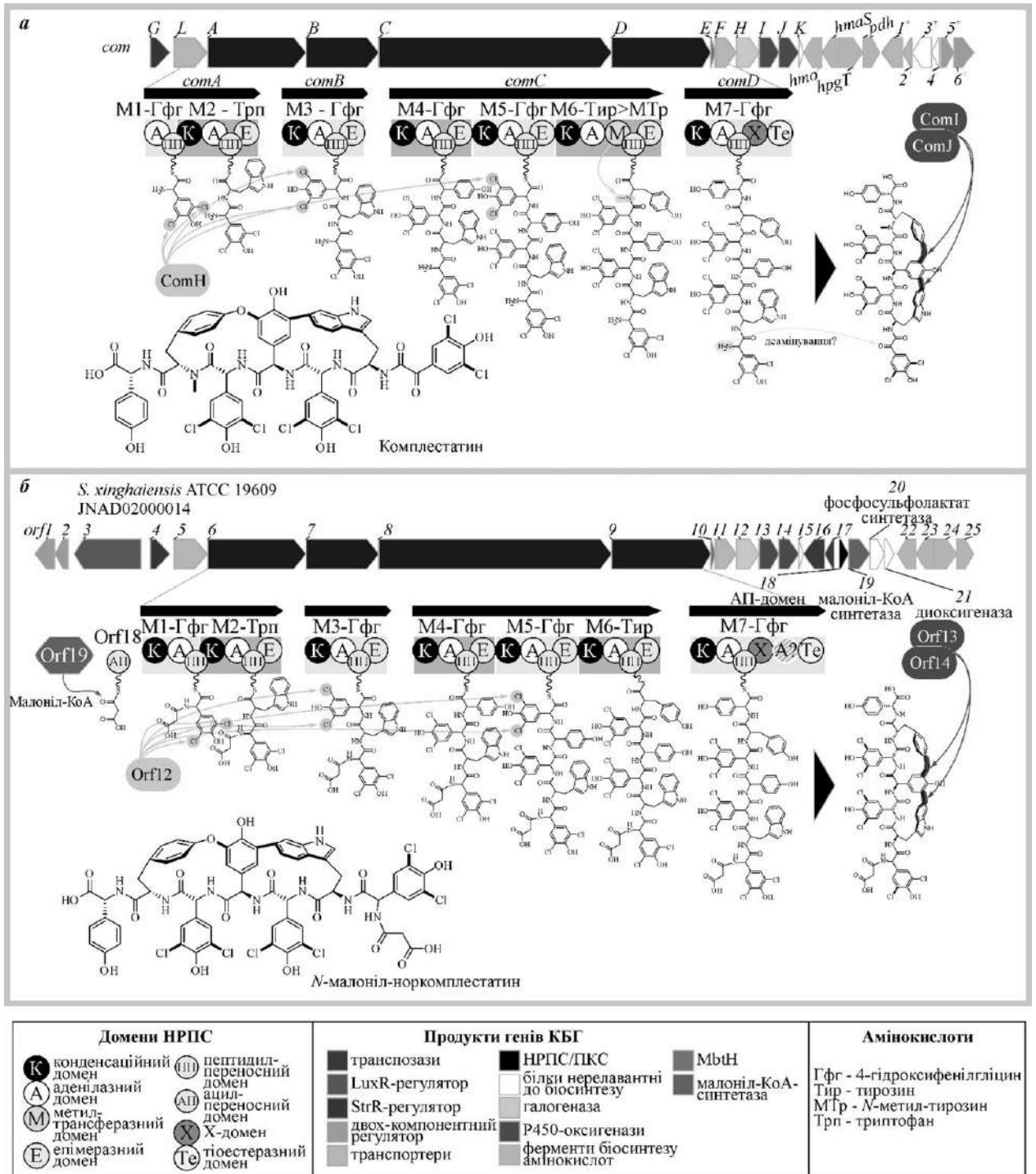


Рис. 2. Інтегровані моделі КБГ та біосинтетичних шляхів комплекстатину (а) і *N*-малоніл-норкомплекстатину (б), що належать до філогенетичних підгруп V та IV КБГ ГПА, подібних до комплекстатину. Умовні позначення розшифровано в легенді до рисунку, деталі обговорюються у тексті статті.

Послідовність *com*-кластера, опублікована раніше [8], містила довгі неанотовані 5'- та 3'-фланги. Анотувавши їх за допомогою antiSMASH, ми виявили в кластері *com* ці ж 6 додаткових генів, які назвали *comI⁺-6⁺* (рис. 2а). Це показує гомогенність структурної організації КБГ підгрупи V, зручність створеного філогенетичного фреймворку і дає змогу зробити важливе доповнення до розуміння організації кластера *com*.

Підгрупа IV. КБГ підгрупи IV (рис. 1) мають організацію, відмінну від *com*-кластера. Зокрема, гени двокомпонентної регуляторної системи містяться не на 3'-кінці кластера, а на його 5'-кінці, а гени двосубодиничного АВС-транспортера відсутні (рис. 2б). Важливо, що у складі КБГ з'явилося сім додаткових генів: ген LuxR-подібного транскрипційного регулятора, два гени транспозаз, ген окремого домена ацил-переносника, гени малоніл-КоА-синтетази, фосфосульфолактонат-синтетази та диоксигенази (рис. 2б). Відмінності спостерігалися і в організації генів НРПС. Зокрема, в гені *comC*, який кодує модуль 4-5-6 комплестатинової НРПС, міститься ділянка, яка кодує унікальний метилтрансферазний домен модуля 6. Функціонування цього домену зумовлює те, що тирозин у шостому положенні гептапептиду комплестатину перебуває у метильованій формі (рис. 2а) [8]. Натомість, ортологи *comC* в кластерах підгрупи IV кодують модуль 6 НРПС, в якому відсутній метилтрансферазний домен. Важливо, що перший модуль НРПС, закодованих в КБГ підгрупи IV, починається із додаткового конденсаційного домена, нагадуючи структуру НРПС іншого ГПА V-го типу – корбоміцину [3]. На кінець, модуль 7 НРПС має у своєму складі додатковий – нефункціональний – аденілазний домен (рис. 2б). Сукупність цих особливостей організації КБГ підгрупи IV дає змогу передбачити структуру ймовірного продукту, який мав би помітно відрізнитися від комплестатину. Зокрема, наявність генів окремого домена ацил-переносника та малоніл-КоА-синтетази, а також присутність додаткового конденсаційного домена в модулі 1 НРПС, дозволяє, за аналогією із корбоміцином [3], припустити, що продукти КБГ підгрупи IV є гібридними НРПС-ПКС молекулами, початковою одиницею у біосинтезі яких виступає малоніл-КоА (рис. 2б). Це, а також відсутність закодованого метилтрансферазного домена, дозволяє говорити про продукт КБГ підгрупи IV як про *N*-малоніл-

норкомплестатин. Подібного роду ГПА досі описано не було, тому цей аспект заслуговує подальшого експериментального вивчення.

Підгрупа III. Ця підгрупа представлена єдиним КБГ із *Streptomyces* sp. BA2 (рис. 1). Цікаво, що він більш подібний до *com*-кластера, ніж до кластерів підгрупи V. Відміни – це розміщення генів двох-компонентної регуляторної системи (ортологів *com5⁺-6⁺*) на 5'-кінці кластера і поява двох додаткових генів: LuxR-регулятора – ортолога генів LuxR-регуляторів КБГ підгрупи IV, і 3-деокси-D-арабіногептулозонат 7-фосфат (ДАГП) синтази (рис. 3а). Гени, подібні до гена цієї синтази, поширені в КБГ ГПА I-IV типів, але раніше не описувалися у складі *com*-подібних кластерів [1]. Цей кластер кодує НРПС, що не містить метилтрансферазного домена в модулі 6. Ці особливості дають підставу припустити, що він мав би кодувати біосинтез норкомплестатину (рис. 3а).

Підгрупа II. Два представники цієї підгрупи, а саме кластери зі *Streptomyces* sp. YIM 121038 і *S. spectabilis* NRRL 2792 (рис. 1), подібні до КБГ підгрупи IV. Вони також кодують біосинтез гіпотетичного *N*-малоніл-норкомплестатину (рис. 3б). Серед відмінностей можна назвати відсутність генів транспозаз, наявність генів двох-компонентного АВС-транспортера, які містяться також у КБГ підгруп V та III, а також певні різниці в орієнтації деяких генів (рис. 3б).

Підгрупа I. КБГ філогенетичної підгрупи I (рис. 1) демонструють ще один різновид організації. У цілому схожі до *com*-кластера, вони не містять ортологів генів *comI⁺-6⁺*, але несуть ортологи генів LuxR-подібного транскрипційного регулятора, присутні також в КБГ підгруп IV-II, а також ДАГП синтази, знайденого в КБГ підгрупи III (рис. 3в). Цікаво, що НРПС, закодовані в КБГ підгрупи I, містять додатковий перший конденсаційний домен в модулі 1, хоча і не мають генів окремого домена ацил-переносника та малоніл-КоА-синтетази, які наявні в КБГ підгруп IV та II. Цей випадок також нагадує ситуацію з біосинтезом корбоміцину [3]. Екстраполюючи його, можна припустити, що біосинтез ГПА, кодований кластерами підгрупи I, потребує для ініціації окремого домена ацил-переносника, що не закодований в КБГ та функціонує *in trans*. Тому, КБГ підгрупи I найімовірніше кодують ще один новий ГПА, а саме *N*-ацетил-норкомплестатин (рис. 3в).

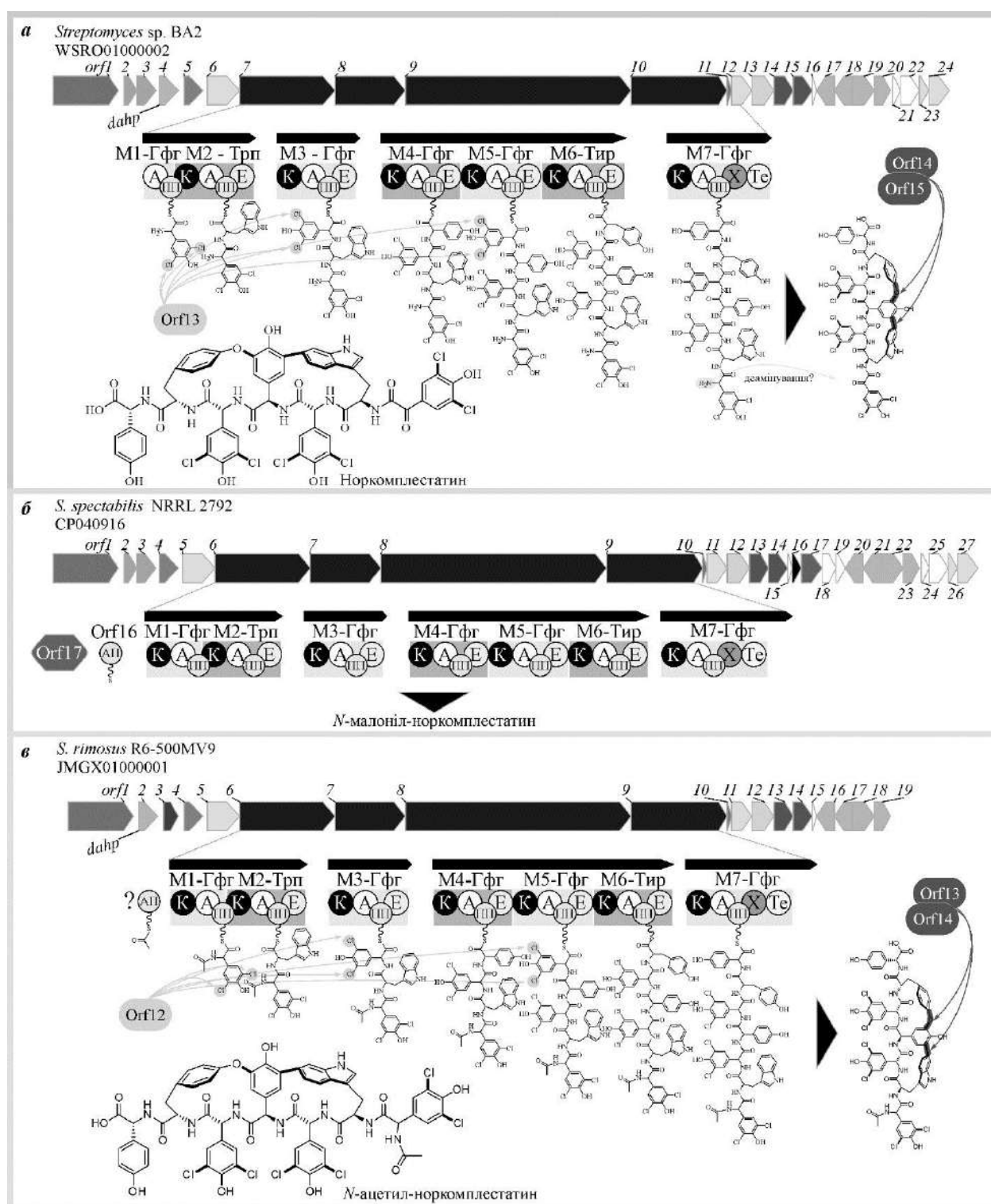


Рис. 3. Інтегровані моделі КБГ та біосинтетичних шляхів норкомплестатину (а), *N*-малоніл-норкомплестатину (б) та *N*-ацетил-норкомплестатину, що належать до філогенетичних підгруп III, II та I КБГ ГПА, подібних до комплестатину. Див. легенду до рис. 2; деталі рисунку обговорюються у тексті статті.

Висновки

У роботі проведено скринінг та аналіз КБГ стрептоміцетів, подібних до кластера біосинтезу комплестатину. Виявлено 33 повних

нових кластерів, які ділять на п'ять філогенетичних підгруп. Кожна з цих підгруп характеризується особливою організацією кластерів. Деякі знахідки заслуговують окремого обговорення.

По-перше, передбачення біосинтетичних шляхів, закодованих в знайдених КБГ, вказує на те, що, ймовірно, існує декілька аналогів комплестатину, а саме норкомплестатин, *N*-малоніл-норкомплестатин і *N*-ацетил-норкомплестатин. Таке теоретичне передбачення спонукає до подальшого експериментального підтвердження продукції цих сполук відповідними штамми. *По-друге*, цікавим є факт наявності гена *LuxR*-подібного транскрипційного регулятора в КБГ усіх підгруп, крім першої, до якої належить і *com*-кластер. Взагалі, основним шлях-специфічним регулятором біосинтезу ГПА є *StrR*-подібний регулятор, що закодований в кожному без винятку КБГ ГПА, у даному випадку це *ComG* та його ортологи (рис. 2, 3). Однак, деякі КБГ ГПА, зокрема тейкопланіну та A40926 [1], містять додатковий ген транскрипційного регулятора *LuxR*-родини. Привабливим виглядає припущення, що *LuxR*-подібні регулятори КБГ підгруп I-IV також беруть участь в регуляції відповідних біосинтетичних шляхів. *По-третє*, порівняння організації КБГ різних підгруп дає змогу висловити припущення про

еволюцію цих КБГ. Зокрема, ймовірно є те, що протокластер підгруп I-IV вже мав ген ДАГП синтази, втрачений у КБГ підгруп II та IV. Також схоже, що гени малоніл-КоА-синтази та окремого домена ацил-переносника потрапили до протокластера підгруп I-IV внаслідок горизонтального генного переносу, про що свідчить наявність генів транспозаз відразу поруч із ним, а пізніше були втрачені в деяких підгрупах. На кінець, наявність унікального метилтрансферазного домена в модулі 6 НРПС, закодованих в КБГ підгрупи V, є, мабуть, унікальним еволюційним здобутком саме цієї підгрупи. У цілому, проведений тут аналіз спонукає нас до подальшого експериментального вивчення комплестатин-подібних КБГ із різних філогенетичних підгруп, що дасть можливість описати нові ГПА типу V.

Електронні додатки. Файли електронних додатків, згаданих у статті (ЕД1, ЕД2, ЕД3), доступні онлайн: <https://data.mendeley.com/datasets/h443z9ztz5/1>, doi: 10.17632/h443z9ztz5.1.

Ця робота була виконана в межах держбюджетної теми Міністерства освіти і науки України БГ-09Ф.

References

1. Yushchuk O., Ostash B. Glycopeptide antibiotics: genetics, chemistry, and new screening approaches. in *Natural Products from Actinomycetes: Diversity, Ecology and Drug Discovery*. 2022. P. 411–444. doi: 10.1007/978-981-16-6132-7_16.
2. Kaneko I., Kamoshida K., Takahashi S. Complestatin, a potent anti-complement substance produced by *Streptomyces lavendulae*. I. Fermentation, isolation and biological characterization. *J. Antibiot. (Tokyo)*. 1989. Vol. 42. P. 236–241. doi: 10.7164/antibiotics.42.236.
3. Culp E.J., Waglechner N., Wang W., Fiebig-Comyn A.A., Hsu Y.P., Koteva K., Sychantha D., Coombes B.K., van Nieuwenhze M.S., Brun Y.V. et al. Evolution-guided discovery of antibiotics that inhibit peptidoglycan remodelling. *Nature*. 2020. Vol. 578. P. 582–587. doi: 10.1038/s41586-020-1990-9.
4. Park Ji-Su, Choi Ha-Young, Kim Won-Gon. The nitrite transporter facilitates biofilm formation via suppression of nitrite reductase and is a new antibiofilm target in *Pseudomonas aeruginosa*. *mBio*. 2020. Vol. 11. P. e00878-20. doi: 10.1128/mBio.00878-20.
5. Momota K., Kaneko I., Kimura S., Mitamura K., Shimada K. Inhibition of human immunodeficiency virus type-1-induced syncytium formation and cytopathicity by complestatin. *Biochem. Biophys. Res. Commun.* 1991. Vol. 179. P. 243–250. doi: 10.1016/0006-291X(91)91361-F.
6. Seo S.Y., Yun B.S., Ryoo I.J., Choi J.S., Joo C.K., Chang S.Y., Chung J.M., Oh S., Gwag B.J., Yoo I.D. Complestatin is a noncompetitive peptide antagonist of N-methyl-D-aspartate and α -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid/kainate receptors: secure blockade of ischemic neuronal death. *J. Pharmacol. Exp. Ther.* 2001. Vol. 299. P. 377–384.
7. Kim E.C., Yun B.S., Ryoo I.J., Min J.K., Won M.H., Lee K.S., Kim Y.M., Yoo I.D., Kwon Y.G. Complestatin prevents apoptotic cell death: inhibition of a mitochondrial caspase pathway through AKT/PKB activation. *Biochem. Biophys. Res. Commun.* 2004. Vol. 313. P. 193–204. doi: 10.1016/j.bbrc.2003.11.104.
8. Chiu H.T., Hubbard B.K., Shah A.N., Eide J., Fredenburg R.A., Walsh C.T., Khosla C. Molecular cloning and sequence analysis of the complestatin biosynthetic gene cluster. *Proc. Natl. Acad. Sci. U. S. A.* 2001. Vol. 98. P. 8548–8553. doi: 10.1073/pnas.151246498.
9. Park O.K., Choi H.Y., Kim G.W., Kim W.G. Generation of new complestatin analogues by heterologous expression of the complestatin biosynthetic gene cluster from *Streptomyces chartreusis* AN1542. *ChemBioChem*. 2016. Vol. 17. P. 1725–1731. doi: 10.1002/cbic.201600241.
10. Waglechner N., McArthur A.G., Wright G.D. Phylogenetic reconciliation reveals the natural history of glycopeptide antibiotic biosynthesis and resistance. *Nat. Microbiol.* 2019. Vol. 4. P. 1862–1871. doi: 10.1038/s41564-019-0531-5.
11. Medema M.H., Takano E., Breitling R. Detecting sequence homology at the gene cluster level with multigeneblast. *Mol. Biol. Evol.* 2013. Vol. 30. P. 1218–1223. doi: 10.1093/molbev/mst025.
12. Blin K., Shaw S., Kloosterman A.M., Charlop-Powers Z., Van Wezel G.P., Medema M.H., Weber T. AntiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res.* 2021. Vol. 49. P. W29–W35. doi: 10.1093/nar/gkab335.

13. Kumar S., Stecher G., Li M., Niyaz C., Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 2018. Vol. 35. P. 1547–1549. doi: 10.1093/molbev/msy096.
14. Jones D.T., Taylor W.R., Thornton J.M. The rapid generation of mutation data matrices from protein sequences. *Bioinformatics* 1992. Vol. 8. P. 275–282. doi: 10.1093/bioinformatics/8.3.275.
15. Kearse M., Moir R., Wilson A., Stones-Havas S., Cheung M., Sturrock S., Buxton S., Cooper A., Markowitz S., Duran C. et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 2012. Vol. 28. P. 1647–1649. doi: 10.1093/bioinformatics/bts199.

YUSHCHUK O.S., ZHUKROVSKA K.A., FEDORENKO V.O.

*Ivan Franko National University of Lviv,
Ukraine, 79005, Lviv, Hrushevskoho str., 4*

DISTRIBUTION OF BIOSYNTHETIC GENE CLUSTERS FOR COMPLESTATIN AND RELATED COMPOUNDS IN *STREPTOMYCES* SPP.

Aim. In current work we have analyzed 2664 genomes of *Streptomyces* spp. (publically available from GenBank) searching for complestatin-like biosynthetic gene clusters (BGCs). Then we aimed to reconstruct *in silico* putative biosynthetic pathways, encoded within these BGCs, estimating the chemical variability of complestatin-like compounds from natural sources. **Methods.** Widely accepted genome-mining tools and approaches for phylogenetic reconstruction were utilized in this work. **Results.** 53 genomes of *Streptomyces* spp. were found to contain complestatin-like BGCs, although only 33 BGCs were found within one contig – others were partial or highly fragmented. Reconstruction of multi-locus phylogeny for 33 found BGCs and complestatin BGC allowed to divide all these BGCs into five phylogenetic subgroups. Representatives of each subgroup exhibited characteristic organization of corresponding BGCs. Reconstruction of putative biosynthetic pathways allowed us to predict that discovered BGCs might potentially code the biosynthesis of new complestatin derivatives: norcomplestatin, *N*-malonyl-norcomplestatin, and *N*-acetyl-norcomplestatin. **Conclusions.** Complestatin-like BGCs are widely distributed among *Streptomyces* spp. and might encode novel complestatin derivatives, which merits further experimental investigation.

Keywords: complestatin, biosynthetic gene clusters, *Streptomyces* spp., genome-mining.