

ЮЩУК О.С.¹, ОСТАШ Б.О.¹, ГОРБАЛЬ Л.О.², ФЕДОРЕНКО В.О.¹¹ Львівський національний університет імені Івана Франка,

Україна, 79005, м. Львів, вул. Грушевського, 4, e-mail: v_fedorenko@franko.lviv.ua

² Helmholtz-Institute for Pharmaceutical Research Saarland,

Germany, Saarbrücken

✉ v_fedorenko@franko.lviv.ua, +38 (032) 239-44-07, +38 (032) 239-44-75

РЕКОНСТРУКЦІЯ ФІЛОГЕНІЇ КЛАСТЕРІВ ГЕНІВ БІОСИНТЕЗУ ГЛІКОПЕПТИДІВ

Клас глікопептидних антибіотиків представлений сполуками, що сьогодні складають останню лінію боротьби проти мультирезистентних Грам-позитивних патогенів [1]. Глікопептиди ванкоміцин та тейкопланін є перевіреними часом клінічними сполуками. Медичну важливість глікопептидних антибіотиків важко переоцінити, що й стимулює інтерес до дослідження цього класу природних сполук. Добре дослідженими глікопептидами, кластери генів біосинтезу яких секвеновано, є лише тейкопланін, ванкоміцин, баліміцин, рістоцетин, хлороеремоміцин, сполуки A40926, A47934 та UK-68597. У базах даних нуклеотидних послідовностей присутні ановані генні кластери, які потенційно можуть кодувати біосинтез глікопептидів. Проте, ці кластери знайдено в секвенованих зразках метагеномної ДНК чи в геномах неідентифікованих бактерій.

Кластери генів біосинтезу глікопептидів (КБГ) досить подібні між собою, що вкладається в модель модульної еволюції кластерів генів біосинтезу природних сполук [2]; продукти окремих генів одних кластерів можуть брати на себе роль своїх гомологів з інших кластерів [3]. Це становить особливий інтерес, якщо врахувати, що штучні комбінації генів біосинтезу глікопептидів можуть потенційно давати нові сполуки. Такий підхід можна застосувати, створюючи нові глікопептиди шляхом комбінаторного біосинтезу. Очевидно, що комбінаторний біосинтез нових глікопептидних сполук ґрунтуються на розумінні філогенії компонентів КБГ у контексті їхніх біологічних функцій. Спроби побудувати філогенію КБГ вже здійснювалися [4], проте, для формування цілісного погляду на цю проблему на той момент було замало секвенованих генів біосинтезу глікопептидів. На сьогодні є значно більше інформації про секвеновані КБГ та більше експериментальних даних. Це дає змогу побудувати ціліснішу картину філогенії

КБГ. У цій роботі ми проаналізували філогенетичні взаємозв'язки окремих компонентів КБГ і на основі отриманих даних запропонували консенсусну модель філогенії цих генних кластерів.

Матеріали і методи

Вихідні нуклеотидні послідовності кластерів біосинтезу глікопептидних антибіотиків отримано із бази даних GenBank; у роботі використано такі послідовності (скорочення, використане в роботі/номер депонування в GB): *tei*/AJ632270 (тейкопланін); UK68597/KF192710 (UK-68,597); *bal*/Y16952 (баліміцин); *van*/HE589771 (ванкоміцин); *risL*, *risJ*/KJ364518, CP008953 (рістоцетин); *cem*/AJ223998-9 (хлороеремоміцин); *VEG*/EU874252; *TEG*/EU874253.1; *A40926*/AJ561198 (*A40926*, попередник дальбаванцину); *A47934*/STU82965 (*A47934*); *CA37*/HM486074; *CA915*/HM486076; *esnapt_15*/KF264554; *pek_WAC1420*, *pek_WAC4229*/JX026280.1, KC688274.1 (пекікоміцини); *com*/AF386507.1 (комплестатин); *feg*/KT809366.1 (фегліміцин). Первинний біоінформатичний аналіз амінокислотних та нуклеотидних послідовностей здійснено за допомогою програми Geneious 4.8.5 [5]. Побудову множинних вирівнювань амінокислотних та нуклеотидних послідовностей, а також конструювання філогенетичних дерев здійснено за допомогою програми MEGA6 [6]. Для побудови множинних вирівнювань використовували MUSCLE із методом ієрархічної кластеризації UPGMA/UPGMV. Дерева у цій роботі створено за допомогою методу Maximum Likelihood, із використанням моделі Джукса-Кантора для нуклеотидних замінів. Топологію фінального дерева будували на основі 500 бутстрепів.

Результати та обговорення

Як вже зазначалося, сьогодні в базах да-

них нуклеотидних послідовностей можна знайти вже багато кластерів, що кодують біосинтез як відомих, так і потенційних глікопептидів. Для проведення філогенетичного аналізу ми відібрали 18 вище перелічених кластерів. Серед «глікопептидних» кластерів присутній і кластер генів біосинтезу лінійного пептидного антибіотика фегліміцину (*feg*), побудованого із 13 амінокислотних залишків (на відміну від 7 таких залишків у глікопептидів), який має багато спільного із «глікопептидними» кластерами. Для обраної вибірки кластерів нами визначено групи ортологічних генів, представники яких присутні у переважній більшості кластерів. Це, зокрема, такі: гени шлях-специфічних регуляторів родини StrR, гени біосинтезу непротеїногенних амінокислот L-3,5-дигідроксифенілгліцину (Dpg) та D-4-гідроксифенілгліцину (Hpg), гени ферментів шикіматного шляху, гени галогеназ і сульфотрансфераз та гени найконсервативніших модулів нерибосомної пептидсинтетази (НРПС4-7). Вони, можливо, були присутні у предковому кластері генів, з якого походять усі сучасні КБГ, і їх можна використати у побудові мультилокусної філогенії останніх. Гени, що присутні лише у деяких кластерах та забезпечують унікальні модифікації (ацилювання, глі-

козилювання тощо), у цій роботі не розглядалися.

Філогенетичні взаємоз'язки генів шлях-специфічних регуляторів біосинтезу глікопептидів. Усі досліджені генні кластери містять ген, що кодує транскрипційний регулятор родини StrR. Експериментально доведено, що StrR-регулятори задіяні в синтезі баліміцину [7], рістоміцину [3], тейкопланіну [8] та A40926 [9]. Усі вони описані як такі, що мають ключове значення для біосинтезу. Ми використали нуклеотидні послідовності відповідних генів для побудови загального філогенетичного дерева 1А (рис. 1 А), використавши як аутгруп послідовність гена *strR* кластера генів біосинтезу стрептоміцину.

На отриманому дереві можна розрізнити дві основні клади, топологія яких була підтверджена максимальними бутстреп-індексами: перша (I) представлена генами з шляхів біосинтезу баліміцину, ванкоміцину, рістоміцинів, пекікоміцинів, A40926 та з кластерів *TEG* та *WEG*. У другу (II) увійшов ген шлях-специфічного регулятора кластера генів біосинтезу фегліміцину. Гени StrR-регуляторів з кластерів біосинтезу A47934 та комплестатину зайняли проміжне положення між двома основними кладами.

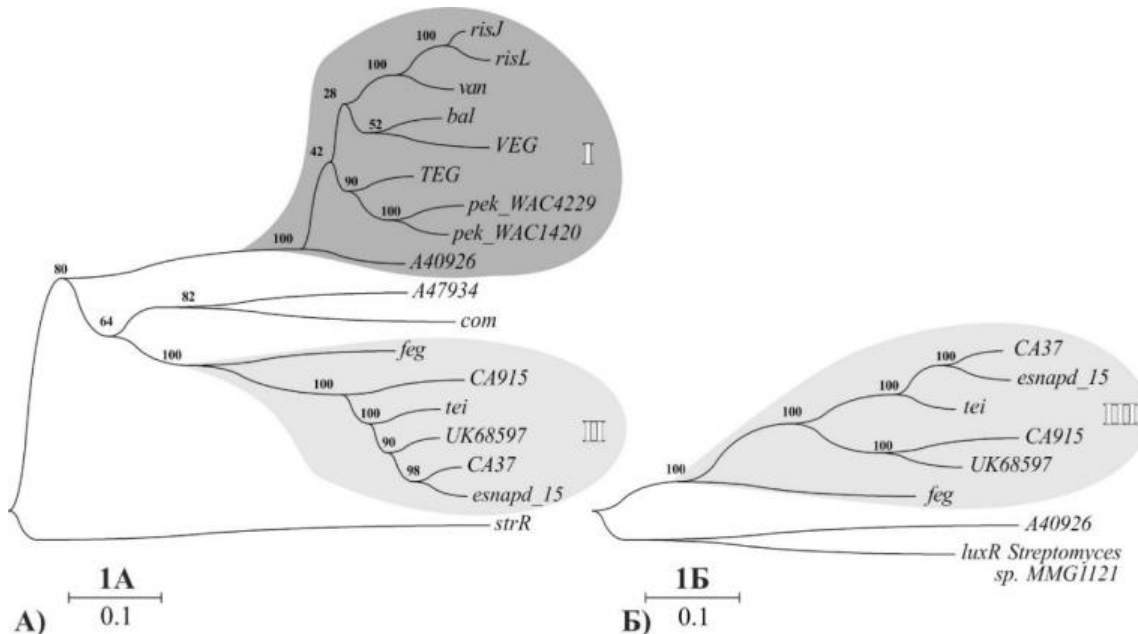


Рис. 1. Філогенетичні дерева, побудовані на основі нуклеотидних послідовностей генів StrR-подібних регуляторів (А) та LuxR-подібних регуляторів (Б) дослідженої вибірки КБГ. Планка масштабу (тут і далі) вказує на кількість нуклеотидних замін на один сайт.

У деяких КБГ присутній другий шлях-специфічний регулятор, що належить до родини LuxR. Вони маловивчені, але відомо, що нокауту генів LuxR-регуляторів блокують біосинтез A40926 та тейкопланіну [8, 9]. Ми спробували відшукати інші LuxR-регулятори у нашій вибірці кластерів. Виявлено ортологи *tei16** (гена LuxR-регулятора з *tei*-кластера) у послідовностях усіх кластерів, що належали до клади II дерева 1А. Отже, StrR-регулятори «глікопептидних» кластерів можна поділити на дві еволюційно віддалені підгрупи. Це корелює із нашими експериментальними даними, які свідчать, що експресія *tei15** в штамі *Amycolatopsis japonicum* не впливала на синтез рістоміцину, тоді як надекспресія гена власного StrR, чи гена StrR з кластера генів біосинтезу баліміцину – активували цей біосинтез [3]. Аналогічно побудовано філогенетичне дерево 1Б (рис. 1 Б) для генів усіх LuxR-регуляторів. Як аутгруп для цього дерева використано ген недослідженого LuxR-регулятора з *Streptomyces* sp. MMG1121. Тут помітно, що всі LuxR-регулятори з кластерів генів біосинтезу тейкопланіну, UK-68,597, фегліміцину і кластерів CA37, CA915, *esnapd_15* формують одну чітко відділену кладу (III). Ген LuxR-регулятора біосинтезу A40926 до неї не входить, проте її аутгрупує разом із геном з *Streptomyces* sp. MMG1121. Цікавою є топологія клади III дерева 1Б, що практично повторює топологію клади II дерева 1А. Імовірно, що спільний предок («протокластер») клад II/III вже мав як StrR-, так і LuxR-регулятор. Водночас, спільний предок клади II мав лише StrR-регулятор, а LuxR-регулятор зі шляху біосинтезу A40926, мабуть, має походження, незалежне від решти досліджених LuxR.

Філогенія генів, що забезпечують біосинтез попередників глікопептидів. До складу агліконів глікопептидних антибіотиків входять непротеїногенні ароматичні амінокислоти, які походять із тирозину. Велику потребу в тирозині як попередникові глікопептидів відображено в структурі генних кластерів: у більшості КБГ є гени, які кодують ферменти шикіматного шляху

біосинтезу ароматичних амінокислот. Це гени префенат-дегідрогеназ (*pdh*) та дезокси-арабіногептулозонат-фосфат-синтаз (*dahp*). Для кластера біосинтезу баліміцину роль таких генів доведено експериментально [10]. Ми виявили, що гени *pdh* присутні у всіх 18 досліджених кластерах, натомість, *dahp* – лише у 12. Для того, щоб оцінити взаємозв'язки між генами *pdh*, ми побудували дерево 2А (рис. 2 А). Як аутгруп ми використали ген *pdh* первинного метаболізму *A. orientalis* (AORI_RS42670).

На дереві 2А можна розрізнити дві головні кладі (IV та VI) та одну мінорну (V). Вміст клад IV та VI збігається із вмістом клад I та II дерева 1А. Аналогічний розподіл спостерігався при аналізі філогенії *dahp* (рис. 2 Б). І тут вміст клад VII та VIII дерева 2Б відповідав вмісту клад IV і VI дерева 2А відповідно.

Філогенія генів, продукти яких задіяні в біосинтезі непротеїногенних амінокислот. До складу агліконів глікопептидних антибіотиків часто входять непротеїногенні амінокислоти, такі як Dpg і Hrg. Кожен кластер містить у своєму складі набори консервативних генів, що контролюють їхній біосинтез. Для філогенетичного аналізу обрано гени *hpgT* та *dpgA*, продукти яких забезпечують біосинтез Hrg та Dpg. Їхні ортологи знайдено у всіх генних кластерах. Винятки – відсутність ортолога *dpgA* у *com*-кластері та TEG-кластері. Втім, гептапептидний попередник комплестатину не містить залишків Dpg, а продукт *teg*-кластера невідомий. Древа ортологів *hpgT* (рис. 3 А) та *dpgA* (рис. 3 Б) мали топологію, характерну і для попередніх дерев (рис. 1, 2). На кожному помітно дві основні кладі, одна з яких містить гени з кластерів *tei*, UK68597, CA37, CA915, *esnapd_15* (клади IX та XII дерев 3А і 3Б), а друга – з генів кластерів *bal*, *van*, *ris*, *pek*, A40926, TEG та VEG (клади X та XI). Ген *hpgT* з *com*-кластера займав проміжне положення між двома основнимикладами дерева 3А, тяжіючи до клади X. Гени *dpgA* з кластерів A47934 та *feg* виявилися близькими до клади XII.

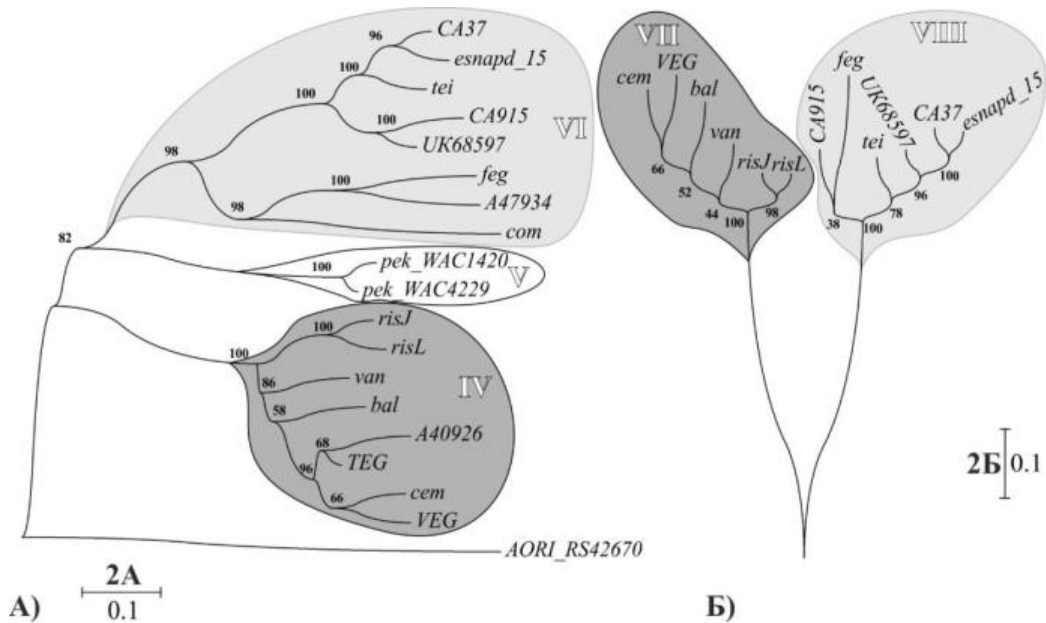


Рис. 2. Філогенетичне дерево, побудоване на основі нуклеотидних послідовностей генів *pdh* (А) та невкорінене дерево генів *dahp* (Б) з дослідженої вибірки КБГ.

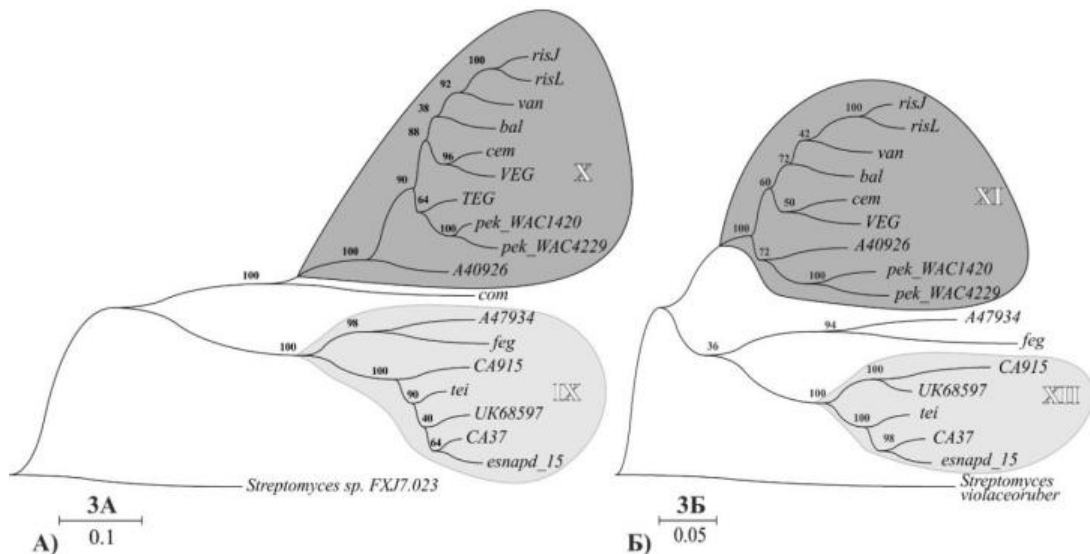


Рис. 3. Деревя, побудовані на основі нуклеотидних послідовностей генів *hpgT* (А) та *dpqA* (Б) дослідженої вибірки КБГ. Як аутгруп для першого було використано послідовність недослідженого гена *hpgT* *S. sp. FXJ7.023*, для другого – *dpqA* кластера генів біосинтезу кендоміцину *S. violaceusruber*.

Філогенія генів, продукти яких задіяні в хлорванні та сульфурілюванні. Серед хімічно вивчених глікопептидів усі, окрім рістоцетинів, аглікони «декоровано» атомами хлору. Найчастіше ним модифіковано 2 і 6 амінокислотні залишки (ак. з.) агліконів (у ванкоміцині, баліміцині, хлорероміцині, тейкопланіні). Крім того, у A47934 хлорованим є 5 ак.з. аглікону, а в UK-68,597 – перший. У A40926 хлоровані 6 і 3 ак.з., а у пекіскоміцині – 6 ак.з. Проду-

центи сполук, що містять атом хлору у положенні, додатковому до стандартних (2 та 6 ак.з.), у своїх генних кластерах містять додатковий ген галогенази. Усі гени галогеназ консервативні, що робить побудову дерев із достовірною топологією неможливою. Єдиною кладою, що підтверджена бутстреп індексом 100, є та, що сформована генами галогеназ кластерів *pek*, *bal*, *VEG* та *van*, що відповідає кладам I, VII, X, XI попередніх дерев. Оскільки кластери генів

біосинтезу рістоцетинів завжди групувалися із генами *pek*, *bal*, *VEG* та *van*, можна припустити, що *ris*-кластери спочатку мали ген галогенази, подібний до генів галогеназ кластерів *pek*, *bal*, *VEG* та *van*, який пізніше втратили. Отже, наймовірніше, первісна галогеназа була здатна включати атом хлору саме в стандартних положеннях. Ми також припускаємо, що гени галогеназ, які приєднують хлор у нестандартних положеннях, виникли унаслідок дуплікації первісного гена та дальшої дивергенції. Сульфурювання глікопептидів відбувається за допомогою сульфотрансфераз, що належать до I надродина. Відповідні гени, знайдені у кластерах *A40926*, *A47934*, *UK68597*, *TEG* та *pek*, консервативні й формують одну щільну кладу на загальному дереві сульфотрансфераз актинобактерій. Експериментальні дані [11] показують, що всі вони мають різну регіоспецифічність дії. Корелювати цю регіоспецифічність із філогенією неможливо.

Філогенетичні взаємозв'язки генів 4–7 модулів НРПС. Біосинтез пептидної частини глікопептидів забезпечується модульними НРПС. Відповідні гени консервативні і мають багато ділянок подібності. Тому різні генні кла-

стери (не тільки всередині родини глікопептидних антибіотиків) здатні обмінюватися окремими генами (їхніми фрагментами), що кодують модулі НРПС. У практично всіх досліджених випадках з нашої вибірки найконсервативнішими є модулі 4–7 НРПС, продукти яких забезпечують включення 4, 5, 6 та 7 ак. з. у гептапептид: двох Hrg, β -гідрокситирозину та Drg відповідно. Винятками є НРПС феєліміцину (принципово інша структура кору) та комплексатину (замість β -гідрокситирозину та Drg включається тирозин та лейцин). Ми побудували філогенетичне дерево на основі генів 4–7 модулів НРПС 17 кластерів нашої вибірки. Топологія отриманого дерева нагадує ту, що спостерігали вище (рис. 1): дві чітко відділені клади, сформовані генами із *tei*, *UK68597*, *CA37*, *CA915*, *esnapd_15* та генами із кластерів *bal*, *van*, *ris*, *pek*, *A40926*, *TEG* та *VEG*.

Підсумовуючи всі дані, нами побудовано мультилокусну філогенію КБГ. Для цього ми конкатенували нуклеотидні послідовності найконсервативніших груп ортологів: *pdh*, *dpgA*, *hpgT*, *strR*-регуляторів та 7 модуля НРПС. Гени *feg*-кластера не враховувалися. Фінальне дерево та коментарі до нього зображено на рис. 4.

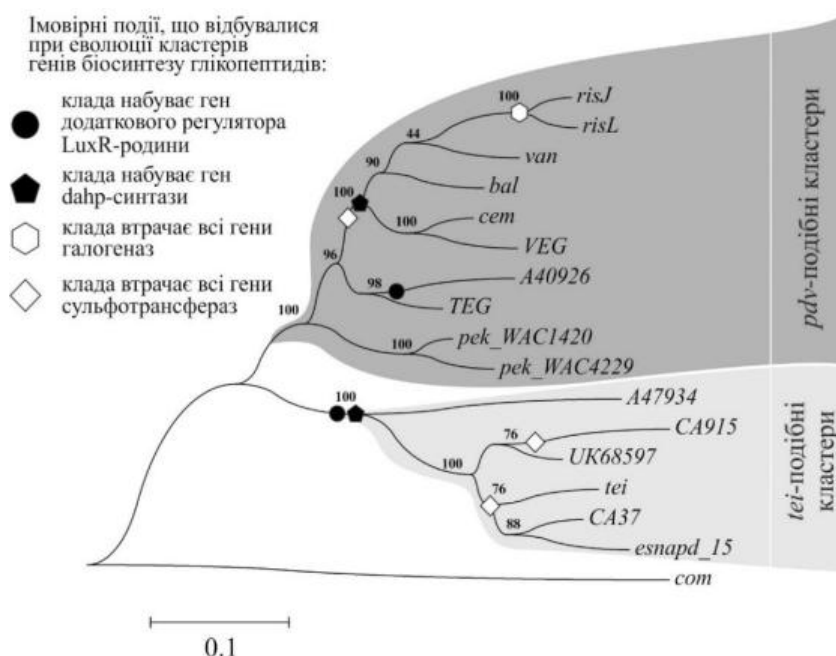


Рис. 4. Філогенетичне дерево, побудоване на основі конкатенатів генів *pdh*, *dpgA*, *hpgT*, *strR* та 7 модуля НРПС вибірки КБГ. На дереві можна розрізнити дві чітко відділені кладу, сформовані *tei*-(тейкопланін)-подібними кластерами та *pdv*-(пекіскоміцин, дальбананцин (*A40926*), ванкоміцин)-подібними кластерами. Перша кладу об'єднує представників I, III, IV структурних типів глікопептидів [12], друга – III та IV. Послідовність генів *com*-кластера (V тип структури) аутгрупує обидві кладу. Легенда описує основні еволюційні події, що вписуються у це дерево.

Висновки

Як на основі мультилокусного порівняння, так і на основі порівнянь окремих консервативних генів, можна стверджувати, що всі кластери генів біосинтезу глікопептидів поділяються на дві основні групи – *tei*-подібні та *pdv*-подібні (рис. 4). Перша об'єднує кластери, які кодують відомі глікопептиди тейкопланін, UK-68,597 та A47934, а також кластери *esnapd_15*, *CA37*, *CA915* із невідомими продуктами. Більшість генів біосинтезу нестандартних сполук – фегліміцину та комплестатину – надзвичайно близькі до першої групи. До другої належать кластери,

що кодують біосинтез рістоцетинів, пекіскоміцинів, ванкоміцину, баліміцину, хлороероміцину, A40926 та кластери VEG і TEG. Філогенетичний розподіл кластерів генів цілком не збігається із структурним розподілом глікопептидів на V типів. Ми припускаємо, що структурна подібність глікопептидів, кластери генів яких філогенетично віддалені, є прикладом конвергентної еволюції. Отримане уявлення про філогенетичні зв'язки КБГ можна використати для планування дальших експериментів із комбінаційного біосинтезу нових сполук.

Література

1. Binda E., Marinelli F., Marcone G.L. Old and new glycopeptide antibiotics: action and resistance // *Antibiotics (Basel)*. – 2014. – V. 3, N 4. – P. 572–594.
2. Medema M., Cimermancic P., Sali A., Takano E., Fischbach M.A. A systematic computational analysis of biosynthetic gene cluster evolution: lessons for engineering biosynthesis // *PLoS Comput. Biol.* – 2014. – V. 10, N 12. – P. 1–12.
3. Spohn M., Kirchner N., Kulik A., Jochim A., Wolf F., Muenzer P., Borst O., Gross H., Wohlleben W., Stegmann E. Overproduction of Ristomycin A by activation of a silent gene cluster in *A. japonicum* MG417-CF17 // *Antimicrob. Ag. Chemother.* – 2014. – V. 58 N 10. – P. 6185–6196.
4. Donadio S., Sosio M., Stegmann E., Weber T., Wohlleben W. Comparative analysis and insights into the evolution of gene clusters for glycopeptide antibiotic biosynthesis // *Mol. Genet. Genomics.* – 2005. – V. 274, N 1. – P. 40–50.
5. Kearse M., Moir R., Wilson A., Stones-Havas S., Cheung M., Sturrock S., Buxton S., Cooper A., Markowitz S., Duran C., Thierer T., Ashton B., Meintjes P., Drummond A. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data // *Bioinformatics.* – 2012. – V. 28, N 12. – P. 1647–1649.
6. Tamura K., Stecher G., Peterson D., Filipinski A., Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0 // *Mol. Biol. Evol.* – 2013. – V. 30, N 12. – P. 2725–2729.
7. Shawky R., Puk O., Wietzorrek A., Pelzer S., Takano E., Wohlleben W., Stegmann E. The border sequence of the balhimycin biosynthesis gene cluster from *A. balhimycina* contains *bbr*, encoding a StrR-like pathway-specific regulator // *J. Mol. Microbiol. Biotechnol.* – 2007. – V. 13, N 1–3. – P. 76–88.
8. Horbal L., Kobylyanskyy A., Truman A.W., Zaburanyi N., Ostash B., Luzhetskyy A., Marinelli F., Fedorenko V. The pathway-specific regulatory genes, *tei15** and *tei16**, are the master switches of teicoplanin production in *A. teichomyceticus* // *Appl. Microbiol. Biotechnol.* – 2014. – V. 98, N 22. – P. 9295–9309.
9. Lo Grasso L., Maffioli S., Sosio M., Bibb M., Puglia A.M., Alduina R. Two master switch regulators trigger A40926 biosynthesis in *Nonomuraea* sp. strain ATCC 39727 // *J. Bacteriol.* – 2015. – V. 197, N 15. – P. 2536–2544.
10. Thykaer J., Nielsen J., Wohlleben W., Weber T., Gutknecht M., Lantz A.E., Stegmann E. Increased glycopeptide production after overexpression of shikimate pathway genes being part of the balhimycin biosynthetic gene cluster // *Metab. Eng.* – 2010. – V. 12, N 5. – P. 455–461.
11. Bischoff D., Bister B., Bertazzo M., Pfeifer V., Stegmann E., Nicholson G.J., Keller S., Pelzer S., Wohlleben W., Süssmuth R.D. The biosynthesis of vancomycin-type glycopeptide antibiotics – a model for oxidative side-chain cross-linking by oxygenases coupled to the action of peptide synthetases // *Chembiochem.* – 2005. – V. 6, N 2. – P. 267–272.
12. Nicolaou K., Boddy C.N., Bräse S., Winssinger N. Chemistry, Biology, and Medicine of the Glycopeptide Antibiotics // *Angew. Chem. Int. Ed. Engl.* – 1999. – V. 38 N 15. – P. 2096–2152.

YUSHCHUK O.¹, OSTASH B.O.¹, HORBAL L.², FEDORENKO V.¹

¹ Ivan Franko National University of Lviv,

Ukraine, 79005, Lviv, Hrushevskoho str., 4, e-mail: v_fedorenko@franko.lviv.ua

² Helmholtz-Institute for Pharmaceutical Research Saarland, Germany, Saarbrücken

RECONSTRUCTING THE PHYLOGENY OF GLYCOPEPTIDE BIOSYNTHETIC GENE CLUSTERS

Aim. In current study we have endeavored to analyze the phylogeny of 18 sequenced glycopeptide biosynthetic gene clusters. **Methods.** Standard methods of phylogenetic reconstruction were used in this research. **Results.** We have created sets of conservative genes, present in all selected clusters and analyzed their phylogeny. In most cases two main clades were formed on final trees. One clade always included genes from biosynthetic clusters of teicoplanin-like com-

pounds, the second – genes from biosynthetic clusters of vancomycin-like compounds. Finally, we have built a multilocus phylogeny (MLP) of selected clusters. Once again, two groups of glycopeptide gene clusters could be recognized according to the results of MLP: biosynthetic clusters of teicoplanin-like compounds and biosynthetic clusters of vancomycin-like compounds. These groups do not correlate with accepted structural classification of glycopeptides. Biosynthetic cluster of unusual antibiotic feglymycin appeared to be closely related to biosynthetic clusters of teicoplanin-like compounds. **Conclusions.** Based on the obtained results we proposed a possible scenario of the evolution of glycopeptide biosynthetic gene clusters.

Keywords: gene cluster, glycopeptide antibiotic, phylogeny.