

АНАЛІЗ БІЛОК-БІЛКОВИХ ВЗАЄМОДІЙ ТИРОЗИЛ-тРНК
СИНТЕТАЗ МЕТОДАМИ БІОІНФОРМАТИКИ

Статтю присвячено теоретичним та практичним аспектам застосування новітніх предиктивних біоінформаційних методів виявлення білок-білкових взаємодій – філогенетичного профілювання та методу генів-сусідів. Названі методи використано в роботі для пошуку ймовірних білків-партнерів тирозил-тРНК синтетаз.

Зважаючи на великі обсяги доступної для аналізу інформації, значного поширення набули біоінформаційні методи досліджень. У статті досліджуються аспекти застосування предиктивних біоінформаційних методів (філогенетичного профілювання та методу генних кластерів) для пошуку ймовірних білкових взаємодій тирозил-тРНК синтетаз.

Основою методу філогенетичного профілювання [1, 5, 9, 11] є визначення профілю досліджуваних білків у сукупності досліджуваних організмів. Філогенетичний профіль — це одновимірна матриця розміром n (де n — кількість досліджуваних організмів), кожен елемент якої може містити 1 або 0 — залежно від наявності (відсутності) гомолога досліджуваного білка в геномі організму з номером елемента. Якщо для двох різних білків філогенетичні профілі ідентичні, то, ймовірно, існує кореляція у схемах еволюційної передачі генів таких білків [1].

Теоретичні основи методу генних кластерів обговорюються у [8, 15]. У прокариотів гени, що входять до кластерів, найчастіше кодують функціонально пов'язані білки [8]. Для еукаріотів також було показано існування окремих оперон-подібних кластерних генетичних структур [11, 12].

Методи

У роботі було використано філогенетичне профілювання [1, 5, 9, 11] та метод генів-сусідів [4, 6, 8, 9, 11]. Отримані результати порівнювалися з даними експериментальних баз даних білок-білкових взаємодій та поточними публікаціями в цьому напрямку [2, 3, 7, 10].

У нашій роботі для пошуку білків-партнерів методом філогенетичного профілювання було використано інструмент філогенетичного пошуку бази даних COG (Clusters of Orthologues Groups) [5, 14]. Параметр наявності білків у організмів було визначено згідно з [5], що забезпечило пошук за всіма 45 організмами.

Аналіз генного оточення для кожного з проаналізованих організмів (всього 21) проводився вручну з використанням генних карт [16] шляхом дослідження безпосереднього оточення генів тирозил-тРНК синтетаз на відповідність установленим критеріям пошуку. Було встановлено такі критерії пошуку для прокариотів і еукаріотів:

- усі гени сукупності мають бути розташовані на одній нитці ДНК,
- відстані між будь-якими двома сусідніми генами не мають перевищувати 300 п. о. [8].

Пошук експериментальних даних про взаємодії тирозил-тРНК синтетаз з білками-партнерами проводився у [17–21] та періодиці [13]. У [17] проводився пошук в розділі Interactions, у базі даних [18] було використано інструмент Text search; аналогічно проводився пошук у [19, 20, 21].

Пошук гомологічних білків з використанням [23] проводився зі стандартними параметрами (матриця BLOSUM62, T = 11, A = 40, X1 = 16, X2 = 38, X3 = 64, S1 = 41, S2 = 69).

Пошук структурних гомологів на основі аналізу Сміта – Уотермана проводився у [24].

Результати та обговорення

У результаті пошуку методом філогенетичного профілювання було отримано 84 групи ортологічних білків (табл. 1).

Отримана кількість груп білків є занадто великою та гетерогенною для висування припущень щодо взаємодій між білками сукупності – зокрема, у [22] та [1] верхньою межею значущості сукупності білків з однаковим профілем є, від-

Таблиця 1. Розподіл груп знайдених білків за функціями

Функціональна група	Кількість білків у функціональній групі
Трансляція, рибосомальні структури та біогенез	55
Транспорт і метаболізм амінокислот	4
Транскрипція	4
ДНК-реплікація, репарація, рекомбінація	4
Транспорт і метаболізм нуклеотидів	4
Посттрансляційна модифікація, шаперони	4
Інші (всього 9 функціональних груп)	16

повідно, 6 і 10 груп. Таким чином, метод філогенетичного профілювання у застосуванні до консервативних систем білків не дає достатньо надійного передбачення білків-партнерів.

Результати аналізу генів-сусідів тирозил-тРНК синтетаз представлено в табл. 2.

Таблиця 2. Кореляція результатів пошуку генів-сусідів з філогенетичним профілюванням

Організми	<i>Drosophila melanogaster</i>	<i>Escherichia coli K12</i>	<i>Bacillus subtilis</i>	<i>Helicobacter pylori 26695</i>	<i>Helicobacter pylori J99</i>	<i>Archaeoglobus fulgidus</i>	<i>Borrelia burgdorferi</i>	<i>Rickettsia prowazekii</i>	<i>Treponema pallidum</i>
Білкі									
Глютамат-тРНК синтетаза	++(*)								
Піридоксаль-кіназа / піридоксин кіназа		++(*)							
Піридоксин-фосфат оксидаза		++(*)							
Рибосомальний білок S4			++(*)						
Уридин-кіназа				++	++				
Фактор ініціації трансляції IF1						++(*)			
Глютаміл-тРНК синтетаза							++		
Фактор ініціації трансляції IF-2								++	
Фактор елонгації транскрипції (білок nusA)								++	
Аргінін-тРНК синтетаза									++
Ацетил-коА синтетаза			+(*)						
Пента-фосфат гуанозин-3' пірофосфогідролаза				+(*)	+(*)				
Субодшинця А АТФ-залежної Clp-протеази							+(*)		
Лікнірин									+(*)

Умовні позначення: + білок відповідає критеріям пошуку генів-сусідів; ++ білок відповідає критеріям пошуку генів-сусідів і входить до результатів філогенетичного профілювання; (*) безпосереднє сусідство з геном тирозил-тРНК синтетази.

Як видно з таблиці, жоден з білків сукупності, отриманої перекириванням результатів обох біоінформаційних методів, не зустрічається в безпосередньому оточенні гена тирозил-тРНК синтетаз більше ніж в одному виді. З цього можемо зробити висновок, що ген тирозил-тРНК синтетаз, імовірно, не входить до еволюційно-сталого кластера генів. Проте для випадків безпосереднього сусідства генів не виключена можливість корельованої експресії генів-сусідів.

Результати пошуку в експериментальних базах даних представлено у табл. 3.

Таблиця 3. Результати пошуку в експериментальних базах даних

Еукаріоти	Прокаріоти
40S рибосомальний білок S9	30S рибосомальний білок S4
YPL013C/MRPS16, мітохондріальний рибосомальний білок (<i>S. cerevisiae</i>)	Псевдоуридин-синтаза С 50S рибосомальної субодиниці
Ймовірний мітохондріальний 40s рибосомальний білок yhr148w (<i>S. Cerevisiae</i>)	Псевдоуридин-синтаза А 30S рибосомальної субодиниці
Ядерний білок Knr4 [14] (<i>S. cerevisiae</i> , регуляція синтезу 1,3-бета-глюкану)	Псевдоуридин-синтаза D 50S рибосомальної субодиниці
	nam9, білок-попередник рибосомального білка S4

Пошук гомологів до білка S4 прокаріотів і білка S9 еукаріотів, проведений у [23], показав значущу гомологію цих білків – до 47 % ідентич-

них залишків (у різних організмів). Гомологію забезпечує домен S4, який входить до обох названих рибосомальних білків, а також до псевдоуридин-синтаз двох родин, до бактеріальної тирозил-тРНК синтетази і до великої кількості малих білків, які беруть участь у процесах регуляції трансляції.

Вивчення структурної гомології мітохондріального рибосомального білка YPL013C/MRPS16 з використанням [24] показало високу (45 % ідентичних, 28 % подібних, $P = 4.8e-15$) структурну гомологію до 30S рибосомального білка S2.

Білок-попередник рибосомального білка S4 був найбільш гомологічним до власне білка S4. Враховуючи те, що серед результатів пошуку для еукаріотів є два мітохондріальних рибосомальних білки, можна очікувати, що для них також буде знайдено прокаріотичні аналоги.

Висновки

Застосування методу генів-сусідів показало, що ген тирозил-тРНК синтетаз з великою ймовірністю не входить до еволюційно-сталого кластера генів. У той же час безпосередні сусіди досліджуваного гена виявили кореляцію з філогенетичним профілюванням.

Пошук в експериментальних базах даних виявив, що лише рибосомальний білок S4 наявний у результатах усіх трьох методів. Порівняння амінокислотного складу дало змогу показати зв'язок між взаємодією тирозил-тРНК синтетаз із білками та наявністю у них S4 домену.

1. *Matteo Pellegrini, Edward M. Marcotte, Michael J. Thompson, David Eisenberg, Todd O. Yeates.* Assigning protein functions by comparative genome analysis: protein phylogenetic profiles // Proc. Natl. Acad. Sci. USA, Biochemistry.– April 1999.– V. 96.– P. 4285–4288.
2. *Gary D. Bader, Ian Donaldson, Cheryl Wolting, B. F. Francis Ouellette, Tony Pawson, Christopher W. V. Hogue.* BIND – the biomolecular interaction network database // Nucleic Acids Research.– 2001.– V. 29, № 1.– P. 242–245.
3. *Gary D. Bader, Christopher W. V. Hogue.* BIND – a data specification for storing and describing biomolecular interactions, molecular complexes and pathways // Bioinformatics.– 2000.– V. 16, № 5.– P. 465–477.
4. *Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, David J. Lipman.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs // Nucleic Acids Research.– 1997.– V. 25, № 17.– P. 3389–3402.
5. *Roman L. Tatusov, Darren A. Natale, Igor V. Garkavtsev, Tatiana A. Tatusova, Uma T. Shankavaram, Bachoti S. Rao, Boris Kiryutin, Michael Y. Galperin, Natalie D. Fedorova, Eugene V. Koonin.* The COG database: new developments in phylogenetic classification of proteins from complete genomes // Nucleic Acids Research.– 2001.– V. 29, № 1.– P. 22–28.
6. *Edward M. Marcotte, Matteo Pellegrini, Ho-Leung Ng, Danny W. Rice, Todd O. Yeates, David Eisenberg.* Detecting protein function and protein-protein interactions from genome sequences // Science.– July 1999.– V. 285.– P. 751–753.
7. *Ioannis Xenarios, Esteben Fernandez, Lukasz Salwinski, Xiaogun Joyce Duan, Michael J. Thompson, Edward M. Marcotte and David Eisenberg.* DIP: the database of interacting proteins: 2001 update // Nucleic Acids Research.– 2001.– V. 29.– P. 239–241.
8. *Ross Overbeek, Michael Fonstein, Mark D'Souza, Gordon D. Pusch, Natalia Maltsev.* The use of gene clusters to infer functional coupling // Proc. Natl. Acad. Sci. USA, Genetics.– March 1999.– V. 96.– P. 2896–2901.
9. *Peer Bork, Thomas Dandekar, Yolande Diaz-Lazcoz, Frank Eisenhaber, Martijn Huynen, Yanping Yuan.* Predicting function: from genes to genomes and back // J. Mol. Biol.– 1998.– V. 283.– P. 707–725.
10. *Takashi Ito, Kosuke Tashiro, Shigeru Muta, Ritsuko Ozawa, Tomoko Chiba, Mayumi Nishizawa, Kiyoshi Yamamoto, Satoru Kuhara, Yoshiyuki Sakaki.* Toward a protein-protein interaction map of the budding yeast: a comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins // PNAS.– February 2000.– V. 97, № 3.– P. 1143–1147.
11. *Insight Progress.*– Nature.– June 2000.– V. 405.– P. 823–826.
12. *Wu, Maniatis T.* A striking organization of a large family of human neural cadherin-like cell adhesion genes // Cell.– June 11, 1999.– 97(6).– P. 779–790.

13. *Dagkessamanskaia A., Martin-Yken H., Basmaji F., Briza P., Francois J.* Interaction of Knr4 protein, a protein involved in cell wall synthesis, with tyrosine tRNA synthetase encoded by TYS1 in *Saccharomyces cerevisiae* // *FEMS Microbiol. Lett.*– June 12, 2001.– 200(1).– P. 53–58.
14. The COG database.– <http://www.ncbi.nlm.nih.gov/COG/>.
15. *Overbeek R., Fonstein M., D'Souza M., Pusch G. D., Maltsev N.* // *In Silico Biol.*– 1998.
16. Genome database.– <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Genome>.
17. BIND – Biomolecular Interaction Network Database.– <http://www.biond.org/>.
18. DIP – Database of Interacting Proteins.– <http://dip.doe-mbi.ucla.edu/>.
19. The GRID.– <http://biodata.mshri.on.ca/grid/index.html>.
20. SGD.– <http://genome-www.stanford.edu/Saccharomyces/>.
21. Protein-Protein Interaction Database.– <http://arrakis.gis.nus.edu.sg/PPiDB/index.html>.
22. *Joseph C. Mellor, Julian Mintseris, Karl H. Clodfelter, Charles DeLisi.* Predictome: a database of putative functional links between proteins.
23. BLAST.– <http://www.ncbi.nlm.nih.gov/BLAST>.
24. Protein DataBase (PDB).– <http://www.rcsb.org/pdb>.

B. Tokovenko, K. Odynets, A. Kornelyuk

ANALYSIS OF PROTEIN-PROTEIN INTERACTIONS OF TYROSYL-TRNA SYNTHETASES USING BIOINFORMATIONAL METHODS

This article discusses theoretical and practical aspects of using modern predictive bioinformational methods of detecting protein-protein interactions – namely phylogenetic profiling and gene neighbours methods. These methods were used in the current study to find the most probable proteins, interacting with tyrosyl-tRNA synthetases.