

Порівняльний аналіз структурної організації генів цитоплазматичної тирозил-тРНК синтетази еукаріотів

К. О. Одинець, О. І. Корнелюк

Інститут молекулярної біології і генетики НАН України
Вул. Академіка Заболотного, 150, Київ, 03143, Україна

E-mail: kornelyuk@imbg.org.ua

Проведено порівняльний біоінформаційний аналіз будови генів цитоплазматичних тирозил-тРНК синтетаз (TyrRS) 54 видів еукаріотів з різних таксонів. В окремих випадках перевірено коректність ідентифікації цих генів та їхніх екзонів шляхом зіставлення з відомими послідовностями кДНК та EST. Порівняння екзонно-інтронної структури виявило, що гени TyrRS хребетних (16 видів) мають значно більшу кількість екзонів (13–14), ніж гени комах (2–7 екзонів), дріжджів і протозоа (1–8 екзонів). У ході еволюції значно збільшується довжина інтронів за рахунок їхнього насичення геномними повторами. Позиції окремих інтронів є консервативними впродовж еволюції генів, що свідчить на користь їхньої ранньої інтеграції у ген спільного попередника хребетних і членистоногих. Для гена TyrRS людини детально проаналізовано альтернативні форми сплайсингу мРНК та геномні повтори, присутні в інтронах.

Ключові слова: тирозил-тРНК синтетаза, гени, екзонно-інтронна структура, еволюція, бази даних.

Вступ. Всебічне розуміння функцій білків у клітині потребує також знання структурної організації і регуляції активності відповідних генів. Типовими представниками білків і генів «домашнього господарства» клітин еукаріотів є цитоплазматичні аміноацил-тРНК синтетази (aaRS), які належать до життєво необхідних компонентів клітини. Завдяки їхній центральній ролі в процесах трансляції генетичного коду aaRS відносять до білків, які дуже рано виникли в ході еволюції. Еукаріоти, як правило, мають повний набір генів для 20 типів цитоплазматичних aaRS та гени, які кодують специфічні aaRS органел (мітохондрій і хлоропластів). Вважається, що aaRS структурних класів I і II виникли в результаті серії генних дуплікацій двох генів-попередників ще до дивергенції ліній про-

каріотів, архебактерій і еукаріотів. Окрім каталітичного домену, який визначає структурний клас та розпізнає і активує відповідну амінокислоту за рахунок ATP, aaRS еукаріотів містять додаткові домени, що обумовлюють специфічність їхньої взаємодії з відповідними тРНК та інші набуті функції, зокрема, додаткові білково-білкові і білково-РНК взаємодії.

Послідовності aaRS та їхніх генів, кількість яких стрімко зростає в доступних банках даних, є об'єктами систематичного аналізу, що має за мету реконструкцію еволюційних подій, таких як рекомбінація, дуплікація, міжвидовий горизонтальний перенос генів тощо [1–3]. Вважається, що гени aaRS взагалі схильні до горизонтального переносу через порівняно слабкі взаємодії aaRS з іншими компонентами клітинних мереж і, отже, вони є зручними маркерами монофілетичності по-

ходження окремих таксонів [4]. Сучасні філогенетичні реконструкції еволюційної історії aaRS, на відміну від попередніх уявлень [5–7], свідчать про її значну складність.

Тирозил-тРНК синтетаза (TyrRS) є гомодимером і каталізує активацію та специфічне приєднання L-тироzinу до відповідних тРНК^{Tyr}. TyrRS є представником родини aaRS класу I (підклас Ic), каталітичний домен яких має структуру так званої згортки Россмана. До цього ж підкласу належить також триптофаніл-тРНК синтетаза. У роботах [3, 8] показано давню дивергенцію генів-попередників цих aaRS. Філогенія TyrRS інтенсивно досліджується [8–11]. Нещодавно виявлено, що еукаріотні цитоплазматичні TyrRS (cytTyrRS) розподіляються на дві основні групи — тварини і гриби, з одного боку, та всі інші еукаріоти — з іншого, з рівнем гомології послідовностей 40–77 % в межах груп та 24–31 % — між групами [12, 13]. У cytTyrRS вищих еукаріотів таксона *Metazoa* з'явився додатковий С-кінцевий модуль розміром приблизно 170 амінокислотних залишків (а. з.), подібний до цитокіну EMAP II [14]. Цей модуль, імовірно, модулює (підсилює) взаємодію TyrRS з тРНК та іншими РНК, тобто є *cis*-кофактором зв'язування тРНК.

У ссавців С-модуль після його вивільнення при протеолітичному розщепленні TyrRS на два окремих модулі виконує (як і білок EMAP II) додаткову сигнальну функцію у міжклітинних взаємодіях, зокрема, при апоптозі [9, 15, 16]. Для N-кінцевого каталітичного модуля TyrRS людини теж виявлено цитокінову активність, подібну до такої інтерлейкіну IL-8 [16].

Завдяки швидким темпам секвенування генів еукаріотних організмів (кількість лише завершених геномних проєктів, за даними Genomes OnLine Database 2.0 (GOLD), становить 40, а розпочатих — 559 [17]) виникла можливість здійснити порівняльний аналіз екзонно-інтронної структури генів їхніх cytTyrRS.

Метою цієї роботи було встановити можливі екзонно-інтронні перебудови та зміни структури генів cytTyrRS у процесі еволюції еукаріотів, локалізувати положення інтронів відносно доменної організації білків і консервативність цих положень, виявити присутність в інтронах геномних повторів з різних родин та альтернативні варіанти сплайсингу мРНК, які можуть виконувати важливу функціональну роль в експресії генів TyrRS.

Матеріали і методи. Аналіз структури генів cytTyrRS проводили для 54 видів еукаріотів та мімівірусу амеби *Acanthamoeba polyphaga* — найбільшого з відомих ДНК вірусів (табл. 1). Зокрема, для хребетних — це вісім видів ссавців, три види інших хордових і чотири види риб. З генів безхребетних аналізували морський огірок *S. purpuratus*, п'ять видів комах і два види круглих хробаків-нематод. З генів інших організмів — це 17 видів грибів, три види рослин, вісім видів *Alveolata* та окремі види з таксонів *Diplomonadida*, *Entamoebidae*, *Euglenozoa* і *Mycetozoa*.

Послідовності генів cytTyrRS разом із фланкуючими ділянками в першу чергу взято з банку даних Entrez Gene з National Center for Biotechnology Information (NCBI) [18], а також в окремих випадках — з депонованих Whole Genome Shotgun (WGS) послідовностей останніх версій геномних проєктів визначених організмів. Послідовності мРНК-транскриптів і білків з банку даних RefSeq [19] знайдено за ключовими словами («tyrosyl-tRNA synthetase», «tyrosine-tRNA ligase», «eukaryota») в разі анотованих послідовностей за допомогою пошукової системи Entrez серверу NCBI. Незалежно проводили пошук за гомологією до перших 360 а. з. cytTyrRS людини з сервісом «protein-protein BLAST» серверу NCBI [20]. Межі генів TyrRS локалізували, зіставляючи геномні послідовності з найдовшими експериментальними послідовностями мРНК/кДНК з бази даних GenBank Nonredundant Division за допомогою сервісу «nucleotide BLAST». Також здійснено множинне вирівнювання відповідних амінокислотних послідовностей TyrRS у ClustalX з наступним ручним уточненням.

Для ідентифікації та візуалізації екзонно-інтронних меж використовували сервери Ensembl, Map Viewer, Sequence Viewer, Evidence Viewer. Альтернативні транскрипти гена TyrRS людини аналізували за допомогою статистичних даних банків даних AceView [21] і UniProt. Для пошуку повторюваних геномних елементів використано сервер RepeatMasker v1.165 [22] і базу даних RepBase Update v.10.04 [23].

Результати і обговорення. Послідовності генів TyrRS взято із загальнодоступних банків даних Entrez Gene та GenBank, при цьому гени мітохондріальних TyrRS вилучали. Всього знайдено гени для 54 видів еукаріотів. За винятком рослин (арабідопсис і рис) та гриба *G. zeae* в геномах дослід-

Таблиця 1
Характеристика генів цитоплазматичних тирозил-тРНК синтетаз та відповідних мРНК і білків для 54 видів еукаріотів

Вид організму	Entrez Gene ID	Кількість екзонів	Хромосома	Розмір генів, п. н.	Код референтної мРНК	Код білка	Кількість вмінокіслотних залишків
<i>Ссавці (Mammalia)</i>							
<i>Homo sapiens</i>	8565	13	1p35.1	42793	NM_003680	NP_003671	528
<i>Pan troglodytes</i>	469266	13/15	1	95735*	XM_524651	XP_524651	763*
<i>Pongo pygmaeus</i>	—	—	—	—	CR859664	CAH91825	528
<i>Macaca fascicularis</i>	—	—	—	—	AB169538	BAE01620	489
<i>Canis familiaris</i>	478148	13/14	2	29984	XM_535324	XP_535324	512
<i>Bos taurus</i>	281581	13	2	32181	NM_174220	NP_776645	528
<i>Mus musculus</i>	107271	13	4D2.2	29656	NM_134151	NP_598912	528
<i>Rattus norvegicus</i>	313047	13	5q36	28653	NM_001025696	NP_001020867	528
<i>Інші хребетні (Chordata)</i>							
<i>Gallus gallus</i>	419666	14	23	4811	NM_001006314	NP_001006314	527
<i>Xenopus laevis</i>	380438	—	—	—	BC045236	AAH45236	528
<i>Xenopus tropicalis</i>	448444	—	—	—	NM_001004987	NP_001004987	528
<i>Giona intestinalis</i>	—	14/16	—	6170	—	ci0100136039	535
<i>Danio rerio</i>	368235	13/14	19	10304	NM_201316	NP_958473	529
<i>Takifugu rubripes</i>	—	—	—	—	—	P00000083615	519
<i>Tetraodon nigroviridis</i>	—	13/14	21	3692	—	CAF99467	531
<i>Комахи (Arthropoda, Insecta) та інші безхребетні</i>							
<i>Anopheles gambiae</i>	1272950	2	2R	1850	XM_311880	XP_311880	542
<i>Apis mellifera</i>	413909	7/9	LG14	3241	XM_623124	XP_623127	522
<i>Drosophila melanogaster</i>	39829	4	3L 72F1	1948	NM_140638	NP_648895	525
<i>Drosophila pseudoobscura</i>	—	—	—	1768	—	EAL30078	525
<i>Drosophila virilis</i>	—	—	—	—	—	AAD03560	237*
<i>Caenorhabditis briggsae</i>	—	—	—	2417	—	CAE72594	497*
<i>Caenorhabditis elegans</i>	173311	4/6	I	10088	NM_170951	NP_740947	512
<i>Strongylocentrotus purpuratus</i>	582945	12	Un	9135	XM_782874	XP_787967	526
<i>Дріжджі (Fungi)</i>							
<i>Ashbya gossypii</i>	—	1	I	1215	—	AAS50218	404
<i>Aspergillus fumigatus</i>	3510945	7	5	1510	XM_748487	XP_753580	391
<i>Aspergillus nidulans</i>	2875830	3	VIII	1291	XM_404194	XP_404194	397
<i>Candida albicans</i>	3638065	1	4	1230	XM_715124	XP_720217	409
<i>Candida glabrata</i>	2888727	1	H	1179	XM_447206	XP_447206	392
<i>Cryptococcus neoformans</i>	3253681	8	1	2921	XM_566587	XP_566587	824
<i>Debaryomyces hansenii</i>	2905286	1	G	1203	XM_462346	XP_462346	400
<i>Encephalitozoon cuniculi</i>	859121	1	V	1014	—	XP_597455	337
<i>Gibberella zeae</i> (gene 1)	2783610	1	1	798	XM_382525	XP_382525	365
<i>Gibberella zeae</i> (gene 2)	2789377	5	4	1423	XM_387612	XP_387612	398

Продовження табл. 1

Вид організму	Entrez Gene ID	Кількість екзонів	Хромосома	Розмір гена, п. н.	Код референтної мРНК	Код білка	Кількість амінокислотних залишків
<i>Дріжджі (Fungi)</i>							
<i>Kluyveromyces lactis</i>	2855997	1	F	1194	XM_455771	XP_455771	397
<i>Magnaporthe grisea</i>	2681546	6	II	1540	XM_365747	XP_365747	391
<i>Neurospora crassa</i>	2710408	5		1492	XP_328461	XM_328460	409
<i>Pneumocystis carinii</i>	—	—	—	—	—	2122232A	370
<i>Saccharomyces cerevisiae</i>	853097	1	VII	1185	L12221	NP_011701	394
<i>Schizosaccharomyces pombe</i>	2539269	1	III	1436	NM_001022868	NP_587876	401
<i>Ustilago maydis</i>	3633965	1	20	1434	XM_756901	XP_761994	477
<i>Yarrowia lipolytica</i>	2909353	1	C	1182	XM_501859	XP_501859	393
<i>Protozoa (Alveolata, Diplomonadida, Entamoebidae, Euglenozoa i Mycetozoa)</i>							
<i>Cryptosporidium hominis</i>	3412420	1	6	1131	XM_662748	XP_667840	376
<i>Cryptosporidium parvum</i>	3375674	1	6	1131	XM_627645	XP_627645	376
<i>Cyanidioschyzon merolae</i>	—	—	—	1260	—	—	420
<i>Dictyostelium discoideum</i>	3394165	3	2	1347	XM_637945	XP_643037	383
<i>Entamoeba histolytica</i>	3404510	1	—	1110	XM_645120	XP_650212	369
<i>Giardia lamblia</i>	3525293	1	—	1164	XM_763449	XP_768542	387
<i>Leishmania major</i>	—	—	—	2049	—	CAJ03211	682
<i>Plasmodium berghei</i>	3426401	3	—	1371	XM_672753	XP_677845	372
<i>Plasmodium chabaudi</i>	3497478	3	—	1374	XM_739271	XP_744364	372
<i>Plasmodium falciparum</i>	2655248	3	8	1547	—	NP_704474	373
<i>Plasmodium yoelii yoelii</i>	3789776	3	—	1368	XM_719359	XP_724451	372
<i>Theileria annulata</i>	—	—	—	1424	—	CA176409	402
<i>Theileria parva</i>	3501161	8	4	1534	XM_758909	XP_764002	427
<i>Trichomonas vaginalis</i>	—	—	—	—	—	—	346
<i>Trypanosoma brucei</i>	3658574	1	7	2093	XM_819156	XP_824249	697
<i>Trypanosoma cruzi</i>	3549441	1	—	2073	XM_812318	XP_817411	690
<i>Рослини (Plants)</i>							
<i>Arabidopsis thaliana</i>	817952	10	2	2532	NM_179891	NP_850222	385
<i>Arabidopsis thaliana</i>	839731	20	1	5317	NM_102601	NP_174157	824
<i>Nicotiana tabacum</i>	—	—	—	—	Y10990	CAA71881	408
<i>Oryza sativa</i>	3046771	1	—	1494	NM_193693	NP_918582	497
<i>Oryza sativa</i>	3049733	5	—	5518	NM_190759	NP_915648	320
<i>Oryza sativa</i>	3069882	9	—	4269	XM_480120	XP_480120	394
<i>Oryza sativa</i>	3070529	10	—	4713	XM_480481	XP_480481	363
<i>Oryza sativa</i>	3072208	7	—	5143	XM_481576	XP_481576	325
<i>Oryza sativa</i>	3067028	9	—	4402	XM_474379	XP_474379	353
<i>Віруси (Viruses)</i>							
Мімівірус з <i>Acanthamoeba polyphaga</i>	3162469	1	—	1041	AAV50399	YP_142478	346

*Імовірно помилкові розміри білкових продуктів, обумовлені нерозпізнанням або додаванням зайвих екзонів.

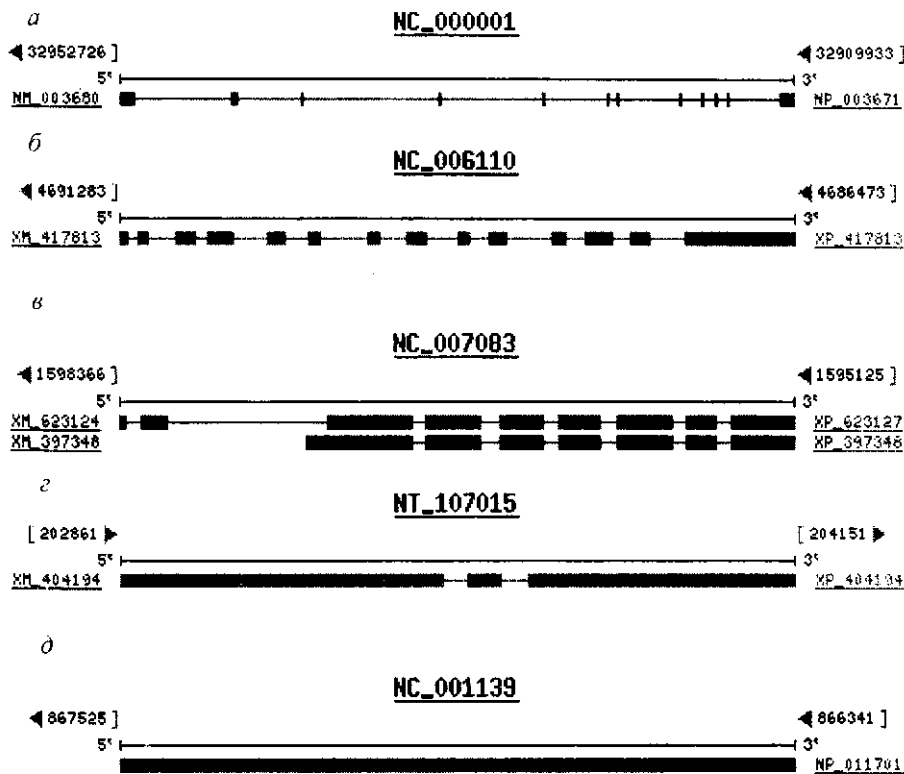


Рис. 1. Порівняння екзонно-інтронної структури генів TygRS для типових представників видів організмів з окремих філогенетичних таксонів, яке демонструє значну варіацію у кількості екзонів та довжинах інтронів: а — людина; б — курка; в — бджола; г — дріжджі *A. nidulans*; д — дріжджі *S. cerevisiae*. Зображення наведено в різних масштабах

жених видів організмів присутній один ген *cut-TygRS*. Ще для семи видів є дані про фрагменти генів або лише про послідовності відповідних мРНК (табл. 1). Задля розширення інформації здійснено також пошук і аналіз доступних послідовностей кДНК інших видів еукаріотів, для яких структура генів ще невідома. Одержано дані для ссавців (орангутанг, макака), патогенних дріжджів (*C. albicans* і *P. carinii*) та рослин (тютюн) (табл. 1). Слід зазначити, що для багатьох видів організмів екзонно-інтронну будову генів передбачено за допомогою автоматизованого комп'ютерного аналізу (наприклад, з використанням програми GNOMON) і не підтверджено порівнянням з експериментальними послідовностями кДНК і EST, отже, в окремих випадках можливі помилки. Для значної кількості генів TygRS анотовано лише кодуючу частину гена (тобто послідовності від старт- до стоп-кодону). При цьому для деяких видів (шимпанзе, собака) з високою імовірністю наведено помилкове передбачення, пов'язане з нерозпізнанням 1-го екзона гена TygRS і додаванням зайвих 5'-кінцевих екзонів. Незалежно було зіставлено відповідні амінокислотні послідовності TygRS за процедурою множинного вирівнювання та перевірено наявність

у них консервативних амінокислотних залишків, що певною мірою підтвердило коректність анотованої екзонно-інтронної структури їхніх генів.

Порівняння екзонно-інтронної структури генів TygRS. Для знайдених генів *cut-TygRS* здійснено порівняльний аналіз екзонно-інтронної структури. Всі проаналізовані екзонно-інтронні границі задовольняють правилу «GT-AG» для меж інтронів та відповідають критеріям, встановленим для цих організмів.

Кількість інтронів у генах досліджених організмів варіює від нуля для різних видів дріжджів та *Alveolata* до 14 у риби *C. intestinalis* (табл. 1, рис. 1). При дослідженні генів TygRS 14 видів хребетних виявлено, що вони мають спільні риси у своїй будові. Ці гени кодують TygRS розміром 519—529 а. з. і містять однакову кількість (13) екзонів (лише у *C. intestinalis* є додатковий внутрішній інтрон), що свідчить про еволюційну стабільність їхньої організації. Положення всіх 12 інтронів відносно кодуючої послідовності і відповідно довжина внутрішніх екзонів у всіх хребетних є однаковими (за винятком кількох випадків незначних зсувів), тобто зберігаються протягом щонайменше 200 млн років. Довжини крайніх ек-

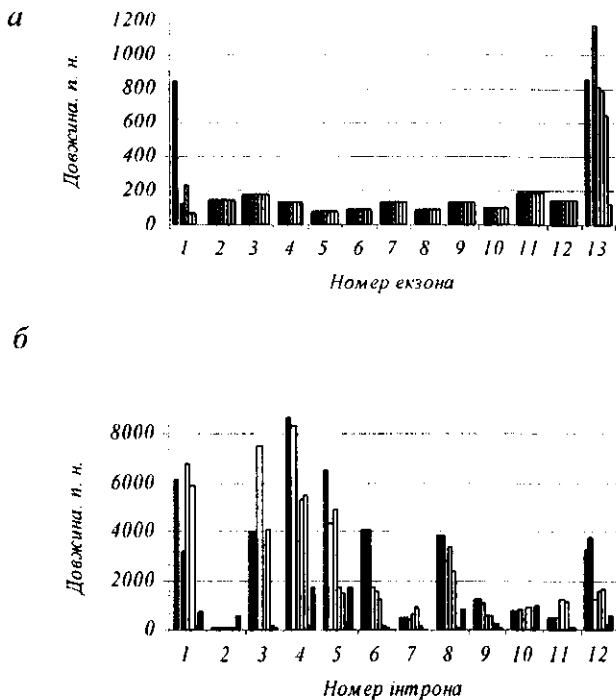


Рис. 2. Порівняння довжин екзонів та інтронів для генів тирозил-тРНК синтетаз хребетних (людина, шимпанзе, собака, миша, щур, курка і риба-зебра). Довжини внутрішніх екзонів майже ідентичні, а довжини інтронів і зовнішніх екзонів значно збільшуються із зростанням еволюційної складності організмів

зонів (E_1 і E_{13}) значно варіюють, при цьому у більшості видів останній екзон має найбільшу довжину, значна частина якого припадає на 3'-нетрансльовану ділянку (3'UTR). Варто відзначити, що сайти початку транскрипції для багатьох генів остаточно не визначені через відсутність статистично достовірних даних про відповідні повнорозмірні транскрипти і, отже, розміри перших екзонів невідомі.

Еволюція інтронів генів *TyrRS*. Порівняння послідовностей інтронів (рис. 2) вказує на те, що їхня середня довжина значно збільшується від риб до людини за рахунок накопичення в них різних видів геномних повторів та дуплікацій під час еволюції. Наприклад, при ідентичних розмірах CDS розмір гена людини на 40 % більший, ніж у миші, а на 12 інтронів припадає 92,7 % довжини гена. При порівнянні з генами риб, які містять дуже короткі інтрони, відмінності ще драматичніші — зокрема, довжина гена риби-фугу складає лише 5,1 % довжини гена людини. Слід зазначити консервативність фази інтронів (тобто їхнє положення відносно нуклеотидів кодонів) у межах так-

сонів. При цьому більшість інтронів (8 з 12) має фазу 0, інтрони I_7 і I_9 мають фазу 1, а інтрони I_3 і I_{11} — фазу 2.

Таким чином, виявлена консервативність структури генів *TyrRS* хребетних свідчить про відсутність екзонно-інтронних перебудов (екзонний shuffling, тобто «стикування» екзонів, втрата екзонів, інтронний слайдинг) під час еволюції хребетних. Це підтверджує, що більша частина інтронів генів *TyrRS* набула спільним попередником хребетних приблизно 200 млн років тому. Однак знайдено істотні відмінності у довжині інтронів генів ссавців і риб (наприклад, сумарна довжина інтронів генів риби-фугу не перевищує 250 п. н.), які можуть свідчити про те, що в процесі еволюції у геномі риби-фугу виробилися захисні механізми стосовно колонізації повторюваними елементами. І, навпаки, у гені *TyrRS* людини повторювані геномні елементи складають до 70 % довжини інтронів (див. далі).

Розташування інтронів у генах *TyrRS* хордових залишається незмінним, що засвідчує спільність їхнього походження і відносно невисокий темп появи нових інтронів. Однак упродовж еволюції в різних таксонах хребетних значно змінюються розміри цих інтронів, перш за все, за рахунок внутрішніх дуплікацій і зростання копійності геномних повторів.

Існує досить чітка кореляція між кількістю інтронів та філогенетичним положенням організму. Так, на відміну від хребетних гени комах мають лише від одного (москіт) до шести (бджола) інтронів. Відповідно розміри генів комах становлять всього 1,7—3,3 тис. п. н. порівняно з 16—33 тис. п. н. у ссавців. Розташування двох інтронів у генах комах корелює з межами основних структурних доменів *TyrRS* хребетних. Положення єдиного інтрона в гені москіта аналогічне до такого інтрона I_2 хребетних. У гені плодової мушки *D. melanogaster* присутні три інтрони, але позиція лише одного з них (I_3) збігається з позицією інтрона I_{11} хребетних. Можна зробити попередній висновок про те, що розміщення цих інтронів збереглося від гена *TyrRS* спільного попередника цих організмів.

Рослини *A. thaliana* і *O. sativa* мають по два гени *cutTyrRS* в районах довгих плечей 1-ї і 2-ї хромосом. Один з них розташований на 1-й хромосомі, відзначається майже удвічі більшою кількістю екзонів (20 екзонів) порівняно з геном *At2g33840* на 2-й хромосомі. Це дає підставу для висновку про

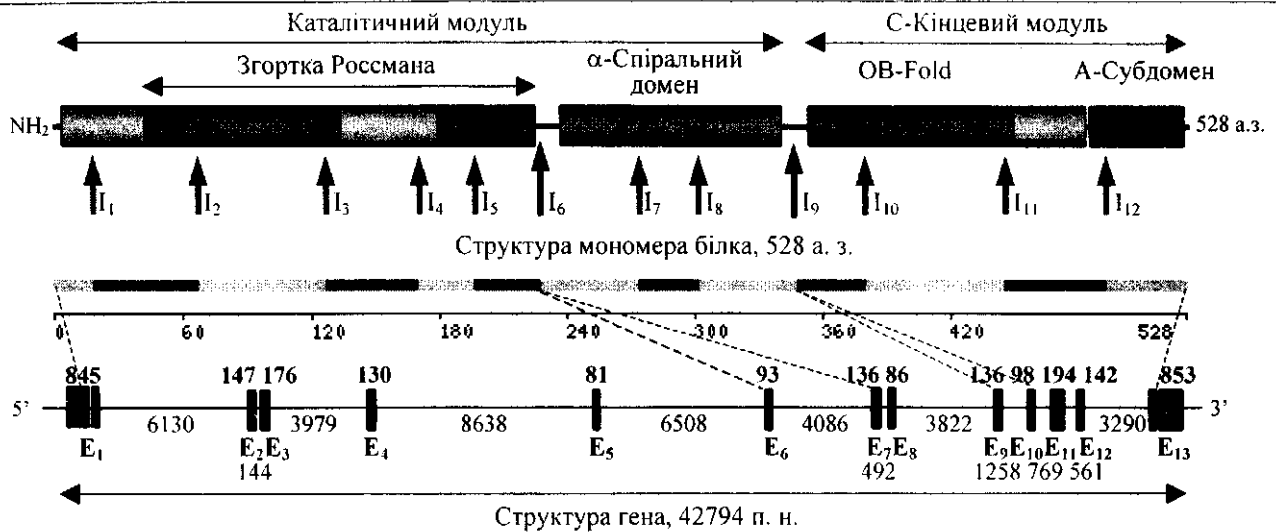


Рис. 3. Зіставлення екзонно-інтронної будови гена цитоплазматичної TygRS людини і доменної організації білка. Показано взаємне розташування 12 інтронів відносно основних структурних доменів TygRS. Розташування інтронів I₆ і I₉ збігається з межами основних структурних доменів TygRS

можливу наявність дуплікації гена TygRS на 1-й хромосомі внаслідок негомологічного кросинговеру або нерівномірного обміну хроматид.

Порівняння екзонної структури генів і доменної структури TygRS еукаріотів. У структурі всіх сугTygRS присутній каталітичний модуль (так звана «мінімальна» форма, mini-TygRS) розміром приблизно 350 а. з., який складається з каталітичного домену з просторовою структурою типу «згортка Россмана» (1—240 а. з., нумерація для гена людини) та антикодон-зв'язувального домену з α-спіральною структурою (241—348 а. з.). У видів організмів таксона *Metazoa* (хребетні і комахи) до каталітичного модуля приєднаний додатковий ЕМАР II-подібний модуль, який складається з «OB fold»-домену (360—468 а. з.) і дистального А-субдомену (469—528 а. з.). Наявність цього модуля у складі TygRS хребетних і комах вказує на давнє набуття цього модуля геном TygRS спільного попередника *Metazoa*. Слід зауважити, що кожний з перелічених доменів є досить автономним у структурному плані і повинен мати свій власний механізм згортання поліпептидного ланцюга. ЕМАР II-подібні модулі присутні також в різних білках апарату трансляції у різних організмів (p43 ссавців, Arc1p дріжджів, MetRS нематод).

У TygRS хребетних найбільший за розміром каталітичний домен (240 з 528 а. з.) кодується екзонами E₁—E₆, а наступний α-спіральний домен (108 а. з.) — екзонами E₇—E₉. С-кінцевий ЕМАР II-подібний модуль кодується екзонами E₁₀—E₁₃, з

яких «OB fold»-домен відповідає екзонам E₁₀, E₁₁ та частково — E₁₂, а А-субдомен (59 а. з.) кодується частково E₁₂ та останнім 13-м екзоном. На екзони E₇, E₁₀ і E₁₂ припадають міждоменні ділянки TygRS хребетних (рис. 3). Таким чином, існує певна кореляція між доменною організацією TygRS та екзонами, які кодують ці домени, що може віддзеркалювати еволюцію генів TygRS.

Екзонно-інтронна структура генів білків з ЕМАР II-подібними модулями. С-кінцевий модуль TygRS є представником родини ЕМАР II-подібних модулів, які мають спорідненість до тРНК, а у ссавців (людина і миша) відповідають за включення апоптозних шляхів у клітині [16]. Зіставлення екзонно-інтронної структури генів TygRS і білка p43 людини показало різне положення їхніх відповідних інтронів (рис. 4, а). Однак в обох цих генах присутні інтрони у міжмодульних лінкерах (I₉ і I₃ відповідно), які з'єднують ЕМАР II-подібні модулі з N-кінцевими модулями згаданих білків. Наявність цих інтронів може вказувати на сайти рекомбінації генів різних N-модулів з геном автономного ЕМАР II-подібного білка.

Дві групи TygRS еукаріотів і гомологія з TygRS археобактерій. Проведене нами порівняння амінокислотних послідовностей сугTygRS еукаріотів вказує на наявність двох філогенетичних груп. До першої групи належать синтетази з *Metazoa* і грибів, до другої — синтетази таксонів *Alveolata*, *Diplomonadida*, *Entamoebidae*, *Euglenozoa*, *Mycetozoa*, зелених рослин, червоних водоростей, а



Рис. 4. Порівняння сайтів сигналів поліаденілювання, розташованих на 3'-кінцях останніх екзонів генів TyrRS хребтних і комах (а) та зіставлення положень кінцевих інтронів відносно амінокислотних послідовностей TyrRS і білка р43 людини (б)

Таблиця 2
Межі і розміри екзонів та інтронів гену TyrRS людини на основі банку даних Entrez Gene (GeneID: 8565)

Екзон			Інтрон				
№	Положення	Довжина, п. н.	№	Положення	Довжина, п. н.	Позиція в блоку	Фаза
1	1—845 (789—845)	845 (57)	1	846—6975	6130	19Q-E	0
2	6976—7122	147	2	7123—7266	144	68E-V	0
3	7267—7442	176	3	7443—11421	3979	127K-E	2
4	11422—11551	130	4	11552—20189	8638	170Q-A	0
5	20190—20270	81	5	20271—26778	6508	197K-Y	0
6	26779—26871	93	6	26872—30957	4086	228E-E	0
7	30958—31093	136	7	31094—31585	492	273S-E	1
8	31586—31671	86	8	31672—35493	3822	302E-V	0
9	35494—35629	136	9	35630—36887	1258	347Q-K	1
10	36888—36985	98	10	36986—37754	769	380K-H	0
11	37755—37948	194	11	37949—38509	561	445I-E	2
12	38510—38651	142	12	38652—41941	3290	492Q-A	0
13	41942—42794 (41942—42052)	853 (111)					

Примітка. У дужках зазначено межі і розміри кодуєчих послідовностей.

Таблиця 3

Основні альтернативні варіанти мРНК гена *TyrRS* людини, які відповідають експериментальним транскриптам цього гена на основі банку даних AceView

Варіант	Кількість клонів	Довжина, п. н.	Кількість екзонів	Довжина транскрипта (пре-мРНК), н.	Положення в гені
aAug05	588	4365	13	44042	1—44042
gAug05	51	2450	5	8328	34146—42473
pAug05	41	2180	2	5470	37443—42912
eAug05	20	957	8	36337	770—37106
dAug05	17	2625	10	22948	19940—42887

також *TyrRS* мімівірусу. Зручним дискримінатором цих груп є мотив, який передує сигнатурі «HIGH» aaRS класу I: у *TyrRS* першої групи — це «TAT(P)T», а другої — «FEPS». Цікаво зазначити, що еукаріотні *TyrRS* другої групи подібні до синте-таз більшості архебактерій (нами проаналізовано *TyrRS* з 28 видів), у яких присутній відповідний мотив «F(Y)EP(I)S», і, отже, їхні гени мають спільне еволюційне походження. Наші висновки збігаються з нещодавно опублікованим розгорну-тим аналізом філогенії *TyrRS* [12, 13].

Регуляторні сигнали в ділянках 5' та 3'UTR і геномні повтори генів TyrRS. Ділянки 3'UTR генів ссавців є досить довгими і становлять ~750 н. (20 % від загальної довжини мРНК), у генів з інших організмів вони значно коротші (50 н.). Гени хребетних містять канонічні сигнали поліадені-лювання AATAAA, розташовані за 15—20 п. н. до polyA послідовностей у відповідних кДНК (рис. 4, б). Проте в генах інших організмів існують варіації цих послідовностей.

Здійснений за допомогою сервісу RepeatMasker пошук геномних повторів та їхніх кластерів для гена *TyrRS* людини показав присутність 97 розта-шованих в інтронах повторюваних елементів з чотирьох родин, значна частина яких є неповними (вкороченими). Зокрема, ідентифіковано 63 SINE елементи (57 Alu і 6 MIR), 18 LINE, 5 LTR і 7 DNA елементів. В інтронах I₁ і I₄ знайдено декілька повноцінних копій LINE елементів, що може вка-зувати на можливість екзонно-інтронного шафлінгу і появи процесованих псевдогенів для гена *TyrRS* у майбутньому.

Характеристика гена TyrRS людини. Для де-тальнішого аналізу було обрано ген цитоплазма-тичної *TyrRS* людини, розташований в локусі р35.1

1-ї хромосоми на (–)-ланцюзі у ділянці 32909—32952 тис. п. н. Цей ген анотовано в банку даних EntrezGene під кодом GeneID:8565 і його послідов-ність виокремлено з послідовності 1-ї хромосоми (GenBank код NC_000001.8), а відповідні рефе-рентні послідовності мРНК і білка — з GenBank (NM_003680.2, розмір 3137 п. н. і NP_003671.1, 528 а. з. відповідно). Ген людини містить 13 ек-зонів (табл. 2, рис. 3) і має GC склад 46,29 %.

Для гена *TyrRS* людини, за даними банку даних UniGene (Hs.213264), відомо 11 повнороз-мірних транскриптів та 874 EST послідовності, а за даними банку даних AceView для версії геному людини «Build 35 version, Aug2005» — 769 експери-ментально виявлених транскриптів, які можна розділити на 17 варіантів, найістотніші з яких наведено в табл. 3 і на рис. 5. Згідно з цією класифікацією, основною формою мРНК є варіант «aAug05», представлений 588 (76 %) транскрипта-ми, який відповідає найдовшій мРНК довжиною 4365 н. Цей варіант мРНК кодується ділянкою гена розміром 44,04 тис. п. н. і містить позну CDS з 13 екзонів, а також має найдовші 5' і 3'UTR (909 і 1865 н. відповідно).

Якщо знехтувати тими альтернативними варі-антами мРНК, які представлені малою кількістю транскриптів (1—4 клони) і можуть бути не пов-ністю процесованими транскриптами або артефак-тами, пов'язаними з їхнім виділенням і секвену-ванням, то, крім основного варіанта, слід від-значити форми «gAug05», «pAug05» і «eAug05» (51, 41 і 20 транскриптів відповідно). Цікаво відмітити, що варіант «gAug05» вміщує останні п'ять екзонів гена, які кодують ЕМАР II-подібний модуль та фрагмент α -спірального домену.

Висновки. Для генів цитоплазматичних *TyrRS*

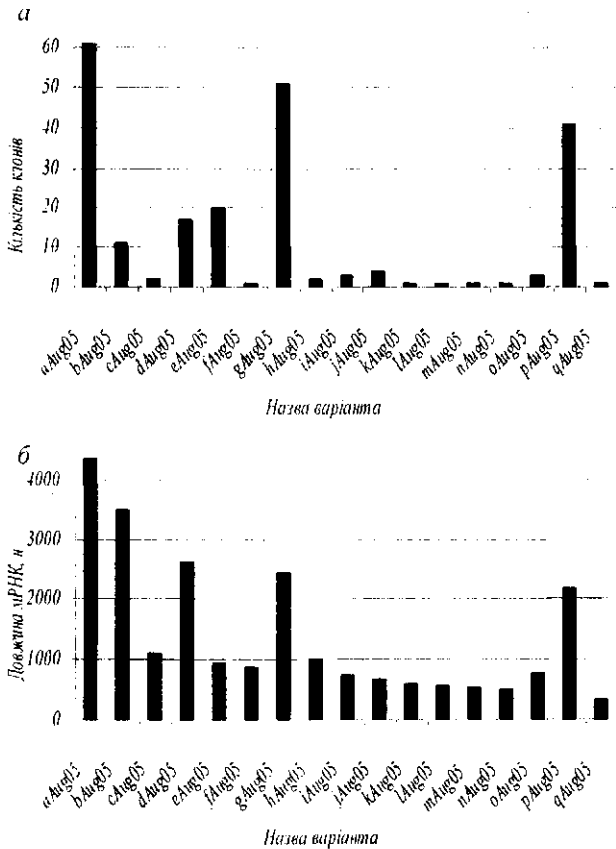


Рис. 5. Порівняння різних варіантів відомих альтернативних транскриптів для гена TyrRS людини за даними банку даних AceView: а — кількість відомих клонів для кожного варіанта; б — довжина транскрипта

з 54 видів еукаріотів проведено аналіз розташування і зміни довжини інтронів протягом еволюції. На прикладі цих ортологічних генів можна в цілому підтвердити відоме положення стосовно того, що гени одноклітинних організмів з великими розмірами популяції містять малу кількість інтронів, яка значно зростає після виникнення багатоклітинних організмів з зменшеними розмірами популяції [24]. Гени TyrRS, представлені, як правило, однією копією на геномі, розташовані у ділянках геному з низьким рівнем рекомбінації, для яких відома тенденція мати більшу кількість і довші інтрони. Розподіл інтронів на користь «фази 0» також має місце для генів TyrRS.

Розмір гена TyrRS значно збільшується в ході еволюції за рахунок зростання кількості і розмірів інтронів. Однак гени хребетних містять інтрони, які практично повністю збігаються за положенням та фазами, тобто ці інтрони існували ще у спільного попередника хордових організмів, а їхня кіль-

кість і положення не змінювалися в процесі еволюції хребетних протягом 200 млн років. Отримані результати також висвітлюють певну залежність між доменною організацією TyrRS та екзонно-інтронною структурою їхніх генів у зазначених еукаріотів. Зібрану анотацію відомих генів цит-ТyrRS сумовано у вигляді спеціалізованої бази даних.

Роботу виконано в рамках програми «Фізіолого-біохімічні та молекулярно-генетичні основи функціонування живих систем і розробка принципів керування ними».

К. А. Одынец, А. І. Корнелюк

Comparative analysis of structural organization of cytoplasmic tyrosyl-tRNA synthetase genes of eukaryotes

Summary

A comparative bioinformational analysis has been performed for the structure of cytoplasmic tyrosyl-tRNA synthetase (TyrRS) genes from 54 eukaryotic organisms from different taxones. In some cases, the identification of the genes and their exons was verified by comparison with the experimental cDNA and EST sequences. The comparison of exon-intron structures shows that TyrRS genes of Chordata (16 species) have much more exons (13–14) than genes of Insecta (2–7 exons), yeasts and Protozoa (1–8 exons). The intron lengths increase intensively during the evolution due to integration of genomic repeats. The positions of some introns are conservative that marks their earlier integration into the gene of common ancestor of Chordata and Insecta. The TyrRS gene of human is analyzed in detail for alternative splicing mRNAs and genomic repeats.

Key words: tyrosyl-tRNA synthetase, gene, exon-intron structure, evolution, databases.

К. А. Одынец, А. И. Корнелюк

Сравнительный анализ структурной организации генов цитоплазматических тирозил-тРНК синтетаз эукариотов

Резюме

Проведен сравнительный биоинформационный анализ строения генов цитоплазматических тирозил-тРНК синтетаз (TyrRS) 54 видов эукариотов из разных таксонов. В отдельных случаях проверена корректность идентификации этих генов и их экзонов путем сопоставления с известными последовательностями КДНК и EST. Сравнение экзонно-интронной структуры показало, что гены TyrRS позвоночных (16 видов) имеют значительно большее количество экзонов (13–14), чем гены насекомых (2–7 экзонов), дрожжей и протозоя (1–8 экзонов), несмотря на эволюционную консервативность доменной организации соответствующих TyrRS. В ходе эволюции значительно увеличивается длина интронов за счет их насыщения геномными повторами. Позиции отдельных интронов сохраняются консервативными в ходе эволюции, что свидетельствует в пользу их ранней интеграции в ген общего предшественника позвоночных и членистоногих. Для гена TyrRS человека детально проанализированы альтернативные формы сплайсинга мРНК и геномные повторы, присутствующие в интронах.

Ключевые слова: тирозил-тРНК синтетаза, гены, экзонно-интронная структура, эволюция, базы данных.

ПЕРЕЛІК ЛІТЕРАТУРИ

1. Berthonneau E., Mirande M. A gene fusion event in the evolution of aminoacyl-tRNA synthetases // FEBS Lett.—2000.—470.—P. 300—304.
2. Brenner S., Corrochano L. M. Translocation events in the evolution of aminoacyl-tRNA synthetases // Proc. Nat. Acad. Sci. USA.—1996.—93.—P. 8485—8489.
3. Diaz-Lazcoz Y., Aude J. C., Nitschke P., Chiapello H., Landes-Devauchelle C., Risler J. L. Evolution of genes, evolution of species: the case of aminoacyl-tRNA synthetases // Mol. Biol. Evol.—1998.—15.—P. 1548—1561.
4. Wolf Y. I., Aravind L., Grishin N. V., Koonin E. V. Evolution of aminoacyl-tRNA synthetases — analysis of unique domain architectures and phylogenetic trees reveals a complex history of horizontal gene transfer events // Genome Res.—1999.—9.—P. 689—710.
5. Doolittle R. F., Handy J. Evolutionary anomalies among the aminoacyl-tRNA synthetases // Curr. Opin. Genet. Develop.—1998.—8.—P. 630—636.
6. Weiner A. M. Molecular evolution: aminoacyl-tRNA synthetases on the loose // Curr. Biol.—1999.—9.—P. R842—R844.
7. Shiba K. Intron positions delineate the evolutionary path of a pervasively appended peptide in five human aminoacyl-tRNA synthetases // J. Mol. Evol.—2002.—55.—P. 727—733.
8. Brown J. R., Robb F. T., Weiss R., Doolittle W. F. Evidence for the early divergence of tryptophanyl- and tyrosyl-tRNA synthetases // J. Mol. Evol.—1997.—45.—P. 9—16.
9. Kleeman T. A., Wei D., Simpson K. L., First E. A. Human tyrosyl-tRNA synthetase shares amino acid sequence homology with a putative cytokine // J. Biol. Chem.—1997.—272.—P. 14420—14425.
10. Chow C. M., RajBhandary U. L. *Saccharomyces cerevisiae* cytoplasmic tyrosyl-tRNA synthetase gene. Isolation by complementation of a mutant *Escherichia coli* suppressor tRNA defective in aminoacylation and sequence analysis // J. Biol. Chem.—1993.—268.—P. 12855—12863.
11. Azad A. K., Stanford D. R., Sarkar S., Hopper A. K. Role of nuclear pools of aminoacyl-tRNA synthetases in tRNA nuclear export // Mol. Biol. Cell.—2001.—12.—P. 1381—1392.
12. Huang J., Xu Y., Gogarten J. P. The presence of a haloarchaeal type tyrosyl-tRNA synthetase marks the opisthokonts as monophyletic // Mol. Biol. Evol.—2005.—22.—P. 2142—2146.
13. Moreira D., Lopez-Garcia P. Comment on «The 1.2-megabase genome sequence of Mimivirus» // Science.—2005.—308.—P. 1114—1115.
14. Kao J., Ryan J., Brett G., Chen J., Shen H., Fan Y. G., Godman G., Familletti P. C., Wang F., Pan Y. C., Stern D., Clauss M. Endothelial monocyte-activating polypeptide II. A novel tumor-derived polypeptide that activates host-response mechanisms // J. Biol. Chem.—1992.—267.—P. 20239—20247.
15. Kornelyuk A. I., Tas M., Dubrovsky A., Murray J. C. Cytokine activity of the non-catalytic EMAP 2-like domain of mammalian tyrosyl-tRNA synthetase // Биополимеры и клетка.—1999.—15.—P. 168—172.
16. Wakasugi K., Schimmel P. Two distinct cytokines released from a human aminoacyl-tRNA synthetase // Science.—1999.—284.—P. 147—151.
17. Bernal A., Ear U., Kyrpides N. Genomes OnLine Database (GOLD): a monitor of genome projects world-wide // Nucl. Acids Res.—2001.—29.—P. 126—127.
18. Maglott D., Ostell J., Pruitt K. D., Tatusova T. Entrez Gene: gene-centered information at NCBI // Nucl. Acids Res.—2005.—33.—P. D54—D58.
19. Pruitt K. D., Tatusova T., Maglott D. R. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins // Nucl. Acids Res.—2005.—33.—P. D501—D504.
20. Altschul S. F., Madden T. L., Scheffer A. A., Zhang J., Zhang Z., Miller W., Lipman D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs // Nucl. Acids Res.—1997.—25.—P. 3389—3402.
21. Lander E. S., Linton L. M., Birren B. Initial sequencing and analysis of the human genome // Nature.—2001.—409.—P. 860—921.
22. Bedell J. A., Korf I., Gish W. MaskerAid: a performance enhancement to RepeatMasker // Bioinformatics.—2000.—16.—P. 1040—1041.
23. Jurka J., Kapitonov V. V., Pavlicek A., Klonowski P., Kohany O., Walichiewicz J. Repbase Update, a database of eukaryotic repetitive elements // Cytogenet. Genome Res.—2005.—110.—P. 462—467.
24. Lynch M. Intron evolution as a population-genetic process // Proc. Nat. Acad. Sci. USA.—2002.—99.—P. 6118—6123.

УДК 577.152.6

Надійшла до редакції 23.12.04