

УДК 575.852.112:577.152.321

## ФУРАНОЗИДАЗНОЕ СУПЕРСЕМЕЙСТВО: ПОИСК ГОМОЛОГОВ

© 2012 г. Д. Г. Наумов<sup>1,2\*</sup>

<sup>1</sup>Институт микробиологии им. С.Н. Виноградского Российской академии наук, Москва, 117312

<sup>2</sup>Государственный научный центр “ГосНИИ генетика”, Москва, 117545

Поступила в редакцию 01.06.2011 г.

Принята к печати 06.07.2011 г.

Фуранозидазное суперсемейство объединяет семейства GH32, GH43, GH62, GH68, GH117, DUF377 и DUF1861 гликозилгидролаз и их гомологов. Каталитические домены этих семейств имеют пространственную структуру пятилопастного  $\beta$ -пропеллера. Итеративный скрининг белковой базы данных позволяет подтвердить их взаимное родство и выявляет эволюционные связи с доменами семейств GH33 и GH93 гликозилгидролаз, имеющих структуру шестилопастного  $\beta$ -пропеллера. Среди гомологов обнаружен также 441 неклассифицированный белок. На основании гомологии эти белки объединены в 39 групп (FURAN1–FURAN39). Группы FURAN8 и FURAN36 могут рассматриваться в качестве самостоятельных подсемейств в составе семейств GH43 и GH32 гликозилгидролаз соответственно, а остальные 37 групп – как новые семейства гипотетических гликозилгидролаз.

**Ключевые слова:** гликозилгидролаза, фуранозидаза, семейства FURAN, новые семейства белков, фуранозидазное суперсемейство, PSI-BLAST, PSI Protein Classifier, CAZy,  $\beta$ -пропеллер, иерархическая классификация белков, эволюция белков, поиск гомологов, аннотация генов.

FURANOSIDASE SUPERFAMILY: SEARCH OF HOMOLOGUES, by D. G. Naumoff<sup>1,2\*</sup> (Winogradsky Institute of Microbiology, Russian Academy of Sciences, Moscow, 117312 Russia; <sup>2</sup>State Research Center “GosNII Genetika”, Moscow, 117545 Russia; \*e-mail: daniil\_naumoff@yahoo.com). The furanosidase superfamily contains GH32, GH43, GH62, GH68, GH117, DUF377, and DUF1861 families of glycoside hydrolases and their homologues. Catalytic domains of these families have five-bladed  $\beta$ -propeller tertiary structure. Iterative screening of the protein database allowed to support their relationship as well as evolutionary connections with domains from GH33 and GH93 families of glycoside hydrolases. The latter two have structure of the six-bladed  $\beta$ -propeller. Among revealed homologues we found 441 unclassified proteins. These proteins are combined into 39 groups based on homology: FURAN1–FURAN39. FURAN8 and FURAN36 can be considered as separate subfamilies within GH43 and GH32 families of glycoside hydrolases, respectively. The remaining 37 groups are new families of hypothetical glycoside hydrolases.

**Keywords:** glycoside hydrolase, furanosidase, FURAN families, new protein families, furanosidase superfamily, PSI-BLAST, PSI Protein Classifier, CAZy,  $\beta$ -propeller, hierarchical protein classification, protein evolution, search of homologues, gene annotation.

“The availability of multiple, essentially complete genome sequences of prokaryotes and eukaryotes spurred both the demand and the opportunity for the construction of an evolutionary classification of genes from these genomes.”

R.L. Tatusov *et al.*, 2003 [1]

Фуранозидазное (или  $\beta$ -фруктозидазное) суперсемейство – это обширная группа гликозилгидролаз (гликозидаз; К.Ф. 3.2.1), отщепляющих от субстрата углеводные остатки, как правило, в фуранозидной форме:  $\beta$ -D-фруктозу,  $\alpha$ -L-арабинозу или их производные. Каталитические домены этих белков имеют пространственную струк-

туру пятилопастного  $\beta$ -пропеллера (five-bladed  $\beta$ -propeller), состоящего из пяти  $\beta$ -слоев, расположенных по кругу в виде тороида. В состав фуранозидазного суперсемейства входят семейства гликозилгидролаз кланов GH-F (семейства GH43 и GH62) и GH-J (GH32 и GH68), а также недавно обособленное из состава GH43 семейство GH117 [2–4]. К этому суперсемейству относят ферменты, расщепляющие гликозидные связи как с сохранением (семейства GH32 и GH68), так и с обращением (GH43) аномальной конфигурации субстрата у продукта реакции гидролиза (retaining и inverting) [3–5]. Активный центр этих гликозидаз содержит три каталитически важных аминокислотных остатка: два Asp и один Glu (рисунок)

\* Эл. почта: daniil\_naumoff@yahoo.com



**Таблица 1.** Представители семейств фуранозидазного суперсемейства, использованные для скрининга базы данных с помощью программы PSI-BLAST

Белок	Семейство	Организм	Размер белка	Область гомологии	Аннотация (NCBI)
3KF3	GH32	<i>Schwanniomyces occidentalis</i>	509	1–331	fructofuranosidase
1UV4	GH43	<i>Bacillus subtilis</i>	293	1–293	arabinanase
CAA03433.1	GH62	<i>Aspergillus niger</i>	295	1–295	arabinofuranosidase
1OYG	GH68	<i>Bacillus subtilis</i>	447	1–447	levansucrase
ABD81917.1	GH117	<i>Saccharophagus degradans</i>	368	1–368	glycosyl hydrolase family 32, N-terminal
1VKD	DUF377	<i>Thermotoga maritima</i>	338	13–338	predicted glycosidase
2B4W	DUF1861	<i>Leishmania major</i>	315	1–315	hypothetical protein

Примечание. В первой колонке указан номер белка в базах данных PDB или GenPept (NCBI). В колонке “Размер белка” указано суммарное число аминокислотных остатков в белке. В колонке “Область гомологии” даны номера аминокислотных остатков в последовательности белка, которые ограничивают участок, использованный в качестве запроса при скрининге базы данных с помощью программы PSI-BLAST. В связи с тем, что в аминокислотной последовательности белка 1VKD содержатся пять неидентифицированных аминокислотных остатков, вместо него при скрининге базы данных использовали белок AAD36300.1.

гидролаз этих двух семейств обнаружить представителей пятидесяти новых для науки семейств гипотетических гликозидаз GH1–GH50, предположительно имеющих каталитические домены со структурой  $(\beta/\alpha)_8$ -бочонка [17–19]. Цель настоящей работы — поиск и классификация энзиматически неохарактеризованных гомологов для белков фуранозидазного суперсемейства.

### АНАЛИЗ ДАННЫХ

Скрининги базы данных аминокислотных последовательностей GenPept на сайте NCBI (<http://www.ncbi.nlm.nih.gov/>) проводили 14–15 апреля и 17 мая 2011 г. при помощи программы PSI-BLAST. В качестве запроса (query) в обоих случаях служили домены семейств GH32, GH43, GH62, GH68, GH117, DUF377 и DUF1861 (табл. 1). Границы доменов определяли на основании экспериментальных данных о пространственной структуре соответствующих белков или их гомологов. По каждому из запросов поиск вели в разделе “non-redundant protein sequences” базы данных GenPept. В качестве порогового значения величины *E*-value для включения найденного белка в модель, используемую программой PSI-BLAST на каждой следующей итерации, использовали 0.005, как это предлагается по умолчанию разработчиками программы. Если с помощью белка-запроса из одного семейства удавалось обнаружить какой-либо белок из другого семейства (с *E*-value  $\leq 0.005$ ), то это называли наличием связи между первым и вторым семейством, а если такая связь обнаруживалась в обоих направлениях, то ее называли двусторонней.

Как и в предыдущих исследованиях [19, 20], критерием для отнесения белка к ранее известному семейству служит обнаружение этого белка с

помощью соответствующего домена какого-либо представителя данного семейства в процессе первой итерации программы PSI-BLAST с *E*-value  $\leq 10^{-7}$ . Белки, не отнесенные ни к одному из известных семейств, объединяли в группы (новые семейства FURAN1–FURAN39) на основании сходства аминокислотных последовательностей доменов. Для отнесения белков к новым семействам применяли тот же численный критерий. В тех случаях, если проявляется достаточный уровень сходства белка сразу с представителями двух семейств, его либо включали в состав того семейства, с представителем которого наблюдается наиболее высокий уровень сходства аминокислотных последовательностей, либо оба семейства объединяли в одно. Для определения принадлежности белков к конкретным семействам (ранее известным или описываемым впервые) применяли программу PSI Protein Classifier [20].

### РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Белковые домены, имеющие пространственную структуру пятилопастного  $\beta$ -пропеллера, применяли для поиска белков-гомологов. По одному домену каждого из семи семейств фуранозидазного суперсемейства (табл. 1) использовали для проведенных дважды итеративных скринингов базы данных аминокислотных последовательностей на сайте NCBI с помощью программы PSI-BLAST. С каждым белком-запросом проводили от 5 до 32 итераций (табл. 2); их число обусловлено техническими ограничениями интернет-версии программы PSI-BLAST. В общей сложности обнаружили 8006 неидентичных белков, в том числе 7552 белка при скрининге в апреле и 7954 белка — в мае 2011 г. Из найденных белков подавляющее большинство содержат в обла-

**Таблица 2.** Эволюционные связи семейств фуранозидазного суперсемейства с другими семействами белковых доменов, по данным скринингов базы данных GenPept программой PSI-BLAST, проведенных 14–15 апреля и 17 мая 2011 г

Семейство белков	GH32 (10/19)	GH43 (5/5)	GH62 (6/8)	GH68 (7/25)	GH117 (6/6)	DUF377 (5/5)	DUF1861 (32/14)	Число белков
GH32	1/1	2/2	3/4	2/2	1/1	2/2	5/5	2509/2588/2597
GH33	—/—	5/—	6/6	—/—	—/—	3/3	7/7	8/7/11
GH43	—/—	1/1	2/2	2/2	1/1	2/2	5/5	3440/3633/3639
GH62	—/—	2/2	1/1	—/—	—/—	4/4	—/—	139/143/143
GH68	—/—	2/2	4/4	1/1	2/2	—/—	—/—	224/303/303
GH93	—/—	—/—	—/—	—/—	—/—	3/3	7/—	10/12/12
GH117	—/—	2/2	4/4	3/3	1/1	2/2	6/6	92/96/96
DUF377	—/—	2/2	4/4	4/—	2/2	1/1	2/2	613/647/647
DUF1861	—/—	—/—	—/—	—/—	—/—	3/3	1/1	99/100/100
FURAN1	—/—	2/2	4/4	4/—	2/2	2/2	5/5	37/38/38
FURAN2	—/—	2/2	4/4	2/2	2/2	2/2	5/5	73/49/76
FURAN3	—/—	—/—	—/—	2/2	—/—	—/—	—/—	3/4/4
FURAN4	8/6	2/2	3/3	—/—	2/2	2/2	5/5	9/9/9
FURAN5	—/—	2/2	4/4	—/—	2/2	2/2	6/6	9/9/9
FURAN6	—/—	—/—	—/—	—/—	3/—	2/2	6/6	27/30/31
FURAN7	—/—	2/2	4/4	—/—	2/2	3/3	—/—	2/2/2
FURAN8	—/—	2/2	3/—	—/—	2/2	3/3	7/—	18/20/20
FURAN9	—/—	2/2	4/—	—/—	2/2	3/3	—/—	4/5/5
FURAN10	—/—	2/2	—/—	—/—	—/—	2/2	7/7	2/2/2
FURAN11	—/—	2/2	6/—	—/—	2/2	2/2	—/7	5/5/5
FURAN12	—/—	2/2	4/4	—/—	—/—	—/—	—/—	4/3/4
FURAN13	—/—	—/—	—/—	—/—	—/—	2/2	6/6	1/1/1
FURAN14	—/—	—/—	—/—	—/—	—/—	3/3	7/7	1/1/1
FURAN15	—/—	4/—	4/4	—/—	—/—	3/3	7/7	11/11/12
FURAN16	—/—	—/—	—/6	—/—	2/2	2/2	7/5	9/11/11
FURAN17	—/—	—/—	4/6	—/—	3/3	2/2	7/7	17/19/19
FURAN18	—/—	—/—	—/—	—/—	—/—	3/3	7/7	1/1/1
FURAN19	—/—	3/3	—/—	—/—	2/2	—/—	—/—	1/1/1
FURAN20	—/—	—/—	6/5	—/—	2/2	2/2	7/7	1/1/1
FURAN21	—/—	3/3	—/6	—/—	—/—	—/—	—/—	1/1/1
FURAN22	—/—	—/—	—/—	—/—	—/—	2/2	8/7	19/25/25
FURAN23	—/—	5/—	—/—	—/—	—/—	3/3	8/—	1/1/1
FURAN24	—/—	—/—	—/—	—/—	2/2	—/—	—/—	4/4/4
FURAN25	—/—	—/—	—/—	—/—	2/2	—/—	—/—	1/1/1
FURAN26	—/—	—/—	—/—	—/—	—/—	3/3	—/—	1/1/1
FURAN27	—/—	—/—	—/—	—/—	—/—	3/3	—/7	1/1/1
FURAN28	—/—	—/—	—/—	—/—	—/—	3/3	—/—	1/1/1
FURAN29	—/—	—/—	—/—	—/—	—/—	3/3	—/—	1/1/1
FURAN30	—/—	—/—	—/—	—/—	—/—	3/3	—/—	3/4/4
FURAN31	—/—	—/—	—/—	—/—	—/—	3/3	—/—	4/5/5
FURAN32	—/—	—/—	—/—	—/—	—/—	3/3	—/—	1/1/1
FURAN33	—/—	—/—	—/—	—/—	—/—	3/3	—/—	1/1/1
FURAN34	—/—	—/—	—/—	—/—	—/—	3/—	—/—	1/0/1
FURAN35	—/—	—/—	6/6	—/—	—/—	—/—	—/—	1/1/1
FURAN36	—/—	2/2	3/3	4/4	1/1	2/2	5/5	134/135/135
FURAN37	—/—	—/—	—/4	—/—	—/—	—/—	—/—	0/2/2
FURAN38	—/—	—/—	—/6	—/—	—/—	—/—	—/—	0/2/2
FURAN39	—/—	—/—	—/—	—/—	—/—	—/3	—/—	0/1/1
COG5263	—/—	4/4	5/5	—/—	—/—	—/—	—/—	8/14/15
PHA03098	—/—	—/—	—/—	—/—	—/—	3/3	—/7	3/4/4
EEN62607.1	—/—	4/—	—/—	—/—	—/—	—/—	—/—	1/0/1
EEZ94632.1	—/—	—/—	—/—	—/—	—/—	—/4	—/—	0/1/1

Примечание. Каждая колонка соответствует одному домену, использованному в качестве запроса (query); список соответствующих белков приведен в табл. 1. В скобках для каждого из запросов указано число итераций, проведенных в апреле и мае 2011 г. В таблице указано минимальное число итераций, необходимое для выявления белка соответствующего семейства (название строки) с  $E\text{-value} \leq 0.005$  по результатам скринингов в апреле и мае. Прочерк означает, что белки этого семейства с соответствующим запросом не обнаружены. В последней колонке указано общее число неидентичных представителей данного семейства, обнаруженных при скрининге хотя бы с одним из семи запросов. Указано число белков каждого семейства, обнаруженных в апреле и мае, и суммарное число неидентичных белков.

сти локального сходства (т.е. предполагаемой гомологии) домен одного из семи семейств фуранозидазного суперсемейства (табл. 2). Еще 23 белка содержат домены семейств GH33 или GH93 гликозилгидролаз (клан GH-E), имеющих структуру шестилопастных  $\beta$ -пропеллеров. У 21-го белка в области сходства обнаружены повторяющиеся последовательности (см. домены семейств COG5263 и PHA03098, а также белки EEN62607.1 и EEZ94632.1 в табл. 2). Это указывает на то, что они не являются, по-видимому, гомологами белков фуранозидазного суперсемейства. Оставшиеся неклассифицированными белки (441 экспериментально неохарактеризованный белок) на основании гомологии объединены нами в 39 групп FURAN1–FURAN39 (табл. 2 и 3). Более подробные исследования показали, что группы FURAN8 и FURAN36 могут рассматриваться в качестве новых подсемейств в составе соответственно семейств GH43 и GH32 на основании наличия “переходных форм” (данные не приводятся). Следует отметить, что из 8006 найденных белков, как минимум, в четырех обнаруживается гомология с двумя различными участками аминокислотной последовательности (т.е. с различными доменами) в зависимости от выбора белка-запроса. Это объясняется тем, что в соответствующих белках имеются два гомологичных домена. Так, белок ACB76517.1 содержит домены GH43 и GH62, ACU62047.1 – домены GH43 и FURAN36, а белки ABR37908.1 и CBK68226.1 – домены GH43 и FURAN8.

Нами обнаружены многочисленные перекрестные (как односторонние, так и двусторонние) связи между семействами фуранозидазного суперсемейства. Например, при помощи каждого из семи белков-запросов в течение первых пяти итераций удалось найти представителя семейства GH32 гликозилгидролаз (табл. 2). Между каждой парой семейств выявлена связь, по крайней мере, в одном направлении; лишь в случае семейств DUF1861 не удалось установить связи с семействами GH62 и GH68 (табл. 2). Это указывает на то, что семейство DUF1861 в фуранозидазном суперсемействе наиболее дивергированное. Обнаружение связей с белками клана GH-E (семейства GH33 и GH93) подтверждает эволюционное родство доменов, имеющих структуру пяти- и шестилопастных  $\beta$ -пропеллеров [15].

При итеративном скрининге базы данных GenPept при помощи аминокислотных последовательностей тахилектина краба (PDB, 1TL2; проведено 4 итерации) и апиразы человека (PDB, 1SID; 3 итерации), не обнаружилось ни одного белка, содержащего домены семейств фуранозидазного суперсемейства или семейств FURAN1–FURAN39 (данные не приводятся). Это указывает либо на независимое происхождение этих двух белков, либо на очень большое эволюционное

расстояние, отделяющее их от остальных доменов, имеющих пространственную структуру пятилопастного  $\beta$ -пропеллера.

Анализ белков, найденных при скрининге базы данных с помощью доменов фуранозидазного суперсемейства, показал, что многие из них состоят из нескольких доменов, что характерно для гликозилгидролаз [4]. В частности, из 441-го гипотетического белка, которые отнесены нами к новым семействам FURAN1–FURAN39, 26 дополнительно содержат домены ряда семейств гликозилгидролаз и гликозилтрансфераз. В одном белке могут одновременно содержаться домены: FURAN1 и GH2 (AAO76103.1, ABR40340.1 и CBK65647.1) или GH106 (ACU59305.1); FURAN2 и GT64 (AAO42055.1, ABF98246.1 и еще 4 белка); FURAN8 и GH43 (ABR37908.1 и CBK68226.1) или GH51 (ACU63160.1, EFB32667.1 и еще 8 белков); FURAN11 и GH28 (EEB24264.1 и EEO47199.1); FURAN22 и GH33 (ZP\_02929332.1); а также FURAN36 и GH43 (ACU62047.1). Следует отметить, что три белка (ACU62047.1, ADB40302.1 и ADZ81571.1), содержащих домен FURAN36, на сайте CAZy отнесены к неклассифицированным гликозилгидролазам, известным также как условное семейство GH0 [5]. Это позволяет предположить, что и остальные белки, содержащие домены FURAN1, FURAN2, FURAN8, FURAN11, FURAN22 или FURAN36, с большой вероятностью являются гликозилгидролазами.

Обнаруженные нами эволюционные связи между доменами белков фуранозидазного суперсемейства и доменами 39 новых групп энзиматически неохарактеризованных белков указывают на то, что домены FURAN1–FURAN39 имеют, по-видимому, пространственную структуру  $\beta$ -пропеллера (наиболее вероятно – пятилопастного) и многие из них могут обладать какими-то гликозилгидролазными активностями (скорее всего фуранозидазными). Для более надежного предсказания числа лопастей в соответствующих  $\beta$ -пропеллерах требуется определить эволюционно наиболее близкий белок с известной пространственной структурой для представителя каждой из этих групп. Группы FURAN8 и FURAN36 могут рассматриваться в качестве самостоятельных подсемейств в составе семейств GH43 и GH32 гликозилгидролаз соответственно, а остальные 37 групп – в качестве новых семейств гипотетических гликозилгидролаз.

Описание новых семейств гипотетических гликозилгидролаз имеет большое значение для аннотации геномных последовательностей, особенно, это касается представителей экспериментально малоизученных таксонов. В качестве примера можно отметить, что гены, кодирующие целый ряд из новых доменов, нам удалось обнаружить у бактерий отделов Acidobacteria (FURAN36 – ABJ88195.1), Planc-

Таблица 3. Белки новых семейств FURAN1–FURAN39 гипотетических гликозилгидролаз

Белок	Семейство	Организм (название отдела или царства)	Размер белка	Область гомологии	Аннотация (NCBI)
AAO76103.1	FURAN1	<i>Bacteroides thetaiotaomicron</i> (Bacteroidetes)	1418	26–428	$\beta$ -galactosidase
AAO42055.1	FURAN2	<i>Arabidopsis thaliana</i> (Viridiplantae)	765	58–332	unknown protein
ABD83216.1	FURAN3	<i>Saccharophagus degradans</i> (Proteobacteria)	373	22–269	hypothetical protein
EEF62904.1	FURAN4	<i>Pedospira parvula</i> (Verrucomicrobia)	297	64–261	hypothetical protein
CAR66584.1	FURAN5	<i>Photorhabdus asymbiotica</i> (Proteobacteria)	514	88–305	conserved hypothetical protein
EDY17437.1	FURAN6	<i>Chthoniobacter flavus</i> (Verrucomicrobia)	542	67–267	hypothetical protein
ACT93517.1	FURAN7	<i>Dyadobacter fermentans</i> (Bacteroidetes)	353	71–292	hypothetical protein
EFB32667.1	FURAN8	<i>Prevotella oris</i> (Bacteroidetes)	887	2–263	$\alpha$ -L-arabino-furanosidase A
ABZ99151.1	FURAN9	<i>Leptospira biflexa</i> (Spirochaetes)	320	17–256	conserved hypothetical protein
EF67960.1	FURAN10	<i>Prevotella</i> sp. (Bacteroidetes)	308	35–269	hypothetical protein
EEO47199.1	FURAN11	<i>Bacteroides dorei</i> (Bacteroidetes)	919	59–354	glycoside hydrolase
EGF25735.1	FURAN12	<i>Rhodopirellula baltica</i> (Planctomycetes)	508	172–391	hypothetical protein
EFB01361.1	FURAN13	<i>Vitivallis vadensis</i> (Lentisphaerae)	296	9–294	hypothetical protein
EDY21898.1	FURAN14	<i>Chthoniobacter flavus</i> (Verrucomicrobia)	322	28–320	hypothetical protein
EDS76506.1	FURAN15	<i>Clostridium botulinum</i> (Firmicutes)	487	176–341	conserved hypothetical protein
ABD11761.1	FURAN16	<i>Frankia</i> sp. (Actinobacteria)	308	25–271	hypothetical protein
EAR10555.1	FURAN17	<i>Reinekea</i> sp. (Proteobacteria)	480	46–246	putative secreted protein
ABD42180.1	FURAN18	<i>Methanospirillum hungatei</i> (Euryarchaeota)	364	93–334	hypothetical protein
ADW14761.1	FURAN19	<i>Geobacter</i> sp. (Proteobacteria)	1361	349–670	collagen triple helix repeat-containing protein
ADV15185.1	FURAN20	<i>Mesorhizobium ciceri</i> (Proteobacteria)	346	27–342	hypothetical protein
ADY57519.1	FURAN21	<i>Syntrophobotulus glycolicus</i> (Firmicutes)	683	50–381	cell wall binding repeat 2-containing protein
EFI39108.1	FURAN22	<i>Bacteroides</i> sp. (Bacteroidetes)	584	103–297	conserved hypothetical protein
ADC64842.1	FURAN23	<i>Ferroplasma placidus</i> (Euryarchaeota)	374	13–233	hypothetical protein
ADB33235.1	FURAN24	<i>Kribbella flavida</i> (Actinobacteria)	1437	1182–1366	carbohydrate binding family 6
EGB08634.1	FURAN25	<i>Aureococcus anophagefferens</i> (stramenopiles)	508	33–164	hypothetical protein
ADE12116.1	FURAN26	<i>Sideroxydans lithotrophicus</i> (Proteobacteria)	309	102–289	hypothetical protein
ADL05430.1	FURAN27	<i>Clostridium saccharolyticum</i> (Firmicutes)	1043	8–253	5'-nucleotidase domain protein
EFR56987.1	FURAN28	<i>Alistipes</i> sp. (Bacteroidetes)	368	137–358	hypothetical protein
EFX59866.1	FURAN29	<i>Daphnia pulex</i> (Metazoa)	322	223–322	hypothetical protein
EFA85763.1	FURAN30	<i>Polysphondylium pallidum</i> (Amoebozoa)	310	104–267	hypothetical protein
EFH03486.1	FURAN31	<i>Methylosinus trichosporium</i> (Proteobacteria)	709	521–636	hypothetical protein
ACL59489.1	FURAN32	<i>Methylbacterium nodulans</i> (Proteobacteria)	386	155–320	conserved hypothetical protein
ACU04840.1	FURAN33	<i>Pedobacter heparinus</i> (Bacteroidetes)	350	58–241	hypothetical protein
EDO56137.1	FURAN34	<i>Bacteroides uniformis</i> (Bacteroidetes)	381	182–370	hypothetical protein
EDP96035.1	FURAN35	<i>Kordia algicida</i> (Bacteroidetes)	696	1–265	hypothetical protein
ABJ88195.1	FURAN36	<i>Solibacter usitatus</i> (Acidobacteria)	402	101–400	glycosyl hydrolase family 32, N-terminal domain protein
BAG26816.1	FURAN37	<i>Lactobacillus fermentum</i> (Firmicutes)	346	60–274	hypothetical protein
EFI04997.1	FURAN38	<i>Bacteroides</i> sp. (Bacteroidetes)	443	150–374	conserved hypothetical protein
EFA74650.1	FURAN39	<i>Polysphondylium pallidum</i> (Amoebozoa)	559	391–520	tyrosinase

Примечание. В первой колонке указан номер соответствующей аминокислотной последовательности белка в базе данных GenPept (NCBI). В колонке "Размер белка" указано суммарное число аминокислотных остатков в белке-предшественнике. В колонке "Область гомологии" даны порядковые номера первого и последнего аминокислотного остатка участка последовательности белка, гомологичного доменам фуранозидазного суперсемейства.

tomycetes (FURAN6 – ADV64171.1 и ZP\_02735497.1; FURAN12 – CAD72109.1 и EGF25735.1) и Verrucomicrobia (FURAN1 – EDY84746.1; FURAN4 – EDY16325.1 и EEF62904.1; FURAN6 – EDY17437.1 и EDY18850.1; FURAN14 – EDY21898.1; FURAN16 – EEF59063.1; FURAN22 – EDY18661.1).

Согласно последней версии базы данных CAZy [5] от 9 ноября 2011 г. семейства GH32, GH43, GH62, GH68 и GH117 содержат 1690, 1883, 90, 157 и 40 белков соответственно.

Работа получила финансовую поддержку Российского фонда фундаментальных исследований (РФФИ 12-04-01155-а).

### СПИСОК ЛИТЕРАТУРЫ

1. Tatusov R.L., Fedorova N.D., Jackson J.D., et al. 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics*. **4**, Art. 41.
2. Наумов Д.Г., Дорошенко В.Г. 1998.  $\beta$ -Фруктозидазы: новое суперсемейство гликозил-гидролаз. *Молекуляр. биология*. **32**, 902–907.
3. Naumoff D.G. 2001.  $\beta$ -Fructosidase superfamily: homology with some  $\alpha$ -L-arabinases and  $\beta$ -D-xylosidases. *Prot. Struct. Funct. Genet.* **42**, 66–76.
4. Наумов Д.Г. 2011. Иерархическая классификация гликозил-гидролаз. *Биохимия*. **76**, 764–780.
5. Carbohydrate-Active Enzymes server. 2011. (<http://www.cazy.org/>).
6. Naumoff D.G. 1999. Conserved sequence motifs in levansucrases and bifunctional  $\beta$ -xylosidases and  $\alpha$ -L-arabinases. *FEBS Lett.* **448**, 177–179.
7. Наумов Д.Г., Лившиц В.А. 2001. Молекулярная структура локуса утилизации сахарозы *Lactobacillus plantarum*: сравнение с *Pediococcus pentosaceus*. *Молекуляр. биология*. **35**, 19–27.
8. Pons T., Naumoff D.G., Martínez-Fleites C., Hernández L. 2004. Three acidic residues are at the active site of a  $\beta$ -propeller architecture in glycoside hydrolase families 32, 43, 62, and 68. *Prot. Struct. Funct. Bioinform.* **54**, 424–432.
9. Naumoff D.G. 2000. Enzymes of fructan synthesis and degradation: sequence analysis and classification. *Abstr. 4<sup>th</sup> Intern. Fructan Symp. "Fructan 2000"*. August 16–20, 2000. Arolla. Switzerland. P. 1.4. (<http://www.kokkinias.com/fructan/admin/abstracts/pdf/naumoff.pdf>).
10. Hashimoto K., Yoshida M., Hasumi K. 2011. Isolation and characterization of CcAbf62A, a GH62  $\alpha$ -L-arabinofuranosidase, from the basidiomycete *Coprinopsis cinerea*. *Biosci. Biotechnol. Biochem.* **75**, 342–345.
11. Rebuffet E., Groisillier A., Thompson A., Jeudy A., Barbeyron T., Czjzek M., Michel G. 2011. Discovery and structural characterization of a novel glycosidase family of marine origin. *Environ. Microbiol.* **13**, 1253–1270.
12. The Pfam database. 2011. Pfam 25.0. (<http://pfam.sanger.ac.uk/>).
13. Murzin A.G., Chandonia J.-M., Andreeva A., Howorth D., Lo Conte L., Ailey B.G., Brenner S.E., Hubbard T.J.P., Chothia C. 2011. Structural classification of proteins. (<http://scop.mrc-lmb.cam.ac.uk/scop/>).
14. Nurizzo D., Turkenburg J.P., Charnock S.J., Roberts S.M., Dodson E.J., McKie V.A., Taylor E.J., Gilbert H.J., Davies G.J. 2002. *Cellvibrio japonicus*  $\alpha$ -L-arabinanase 43A has a novel five-blade  $\beta$ -propeller fold. *Nat. Struct. Biol.* **9**, 665–668.
15. Chaudhuri I., Söding J., Lupas A.N. 2008. Evolution of the  $\beta$ -propeller fold. *Prot. Struct. Funct. Bioinform.* **71**, 795–803.
16. Naumoff D.G. 2006. Development of a hierarchical classification of the TIM-barrel type glycoside hydrolases. *Proc. Fifth Intern. Conf. Bioinform. Genome Regulat. Struct.* July 16–22, 2006. Novosibirsk. Russia. **1**, 294–298. ([http://www.bionet.nsc.ru/meeting/bgrs\\_proceedings/papers/2006/BGRS\\_2006\\_V1\\_067.pdf](http://www.bionet.nsc.ru/meeting/bgrs_proceedings/papers/2006/BGRS_2006_V1_067.pdf)).
17. Naumoff D.G. 2010. GH101 family of glycoside hydrolases: subfamily structure and evolutionary connections with other families. *J. Bioinform. Comput. Biol.* **8**, 437–451.
18. Наумов Д.Г., Степушенко О.О. 2011. Эндо- $\alpha$ -1,4-полигалактозаминидазы и их гомологи: структура и эволюция. *Молекуляр. биология*. **45**, 703–714.
19. Наумов Д.Г. 2011. GHL1–GHL15: новые семейства гипотетических гликозилгидролаз. *Молекуляр. биология*. **45**, 1073–1083.
20. Наумов Д.Г., Карперас М. 2009. Новая программа PSI Protein Classifier автоматизирует анализ результатов программы PSI-BLAST. *Молекуляр. биология*. **43**, 709–721.