

УДК 579.222.7: 579.252.2

БИОИНФОРМАТИЧЕСКИЙ АНАЛИЗ ГЕНОМА ШТАММА *GEOBACILLUS STEAROTHERMOPHILUS* 22, ВЫДЕЛЕННОГО ИЗ ГОРЯЧЕГО ИСТОЧНИКА ГАРГА (ПРИБАЙКАЛЬЕ)

© 2013 г. А.С. Розанов¹, Т.В. Иванисенко¹, А.В. Брянская¹,
С.В. Шеховцов¹, М.Д. Логачева², О.В. Сайк¹, Т.К. Малуп¹,
П.С. Деменков¹, Т.Н. Горячкова¹, В.А. Иванисенко¹, С.Е. Пельтек¹

¹ Федеральное государственное бюджетное учреждение науки Институт цитологии и генетики
Сибирского отделения Российской академии наук, Новосибирск, Россия,
e-mail: peltek@bionet.nsc.ru;

² Московский государственный университет им. М.В. Ломоносова, Москва, Россия

Поступила в редакцию 15 августа 2013 г. Принята к публикации 5 сентября 2013 г.

Новый штамм *Geobacillus stearothermophilus* 22 был выделен из термального источника Гарга, расположенного в Баргузинской долине Прибайкалья. Были проанализированы морфологические и биохимические особенности *G. stearothermophilus* 22, проведено полногеномное секвенирование с последующим биоинформатическим анализом. Показана высокая степень сходства нуклеотидных последовательностей (контигов) анализируемого штамма с геномом бактерии-термофила *G. kaustophilus* Y412MC52. Охарактеризован протеом выделенной бактерии. Обнаружены ферменты, относящиеся к гемицеллюлазам (эндоксилаза, бета-ксилозидаза, арабинофуранозидаза), и фермент эндоксилаза.

Ключевые слова: *Geobacillus*, термофилы, полногеномное секвенирование, биоинформатический анализ.

ВВЕДЕНИЕ

Микроорганизмы, обитающие в высокотемпературных условиях, представляют огромный интерес с точки зрения получения термостабильных ферментов, а также как потенциальные клеточные катализаторы. Для разработки клеточных катализаторов перспективными являются бактерии рода *Geobacillus*, который включает широкий спектр термофильных микроорганизмов с различной физиологией. Известно, что представители рода *Geobacillus* являются хемоорганотрофами, аэробами или факультативными анаэробами, термофилами, имеющими температурный диапазон роста 40–75 °С с оптимумом 55–65 °С, диапазон pH 6,0–8,5 с оптимумом 6,2–7,5. Данные микроорганизмы представляют интерес ввиду их высокой скорости роста и способности утилизировать широкий круг субстратов, включая

пентасахара. Кроме того, эти микроорганизмы обладают уникальными гемицеллюлолитическими системами, что позволяет рассматривать их в качестве потенциальных источников высокоактивных и термостабильных ферментов для эффективного гидролиза биомассы (Brook *et al.*, 1978; Бонч-Осмоловская и др., 2004).

В настоящее время опубликовано несколько работ, в которых было выполнено изменение геномов для улучшения целевых свойств микроорганизмов (Taylor *et al.*, 2008). В том числе было показано, что возможна трансформация ранее нетрансформируемых или очень плохо трансформируемых бактерий после использования модифицированной ДНК, в соответствии с системой рестрикции-модификации штамма-реципиента (Suzuki, Yoshida, 2012). В этой работе было наглядно продемонстрировано, как знания о геноме бактерии могут помочь в разработке методики модификации генома

микроорганизма. Полногеномный анализ необходим в силу того, что генетические последовательности бактерий, выделенных из различных источников, могут значительно различаться по составу генов, несмотря на близость по стандартным филогенетическим маркерам. Кроме того, знание последовательности генома позволяет направленно модифицировать гены для получения штаммов-продуцентов с заданными свойствами (Cripps *et al.*, 2009).

Целью данной работы являлось описание штамма *G. stearothermophilus* 22, его полногеномное секвенирование и биоинформатический анализ полученных данных для выявления особенностей катаболизма и определения генетических и белковых последовательностей, кодируемых в геноме. В настоящее время доступны полные геномные последовательности двух видов, относящихся к роду *Geobacillus*: *G. kaustophilus* и *G. thermodenitrificans*, а также частично расшифрованы геномы некоторых штаммов, относящихся к виду *G. stearothermophilus*.

С помощью биоинформатического анализа вновь секвенированного генома нами была установлена наибольшая степень его сходства среди всех аннотированных геномов с геномом *G. kaustophilus* Y412MC52. В геноме вновь секвенированной бактерии были идентифицированы все гены, ответственные за гликолитический метаболизм, включая гены лактат дегидрогеназы, ацетальдегид дегидрогеназы, алкоголь дегидрогеназы, ацетат киназы и пироват дегидрогеназы.

МАТЕРИАЛЫ И МЕТОДЫ

Штамм *G. stearothermophilus* 22 был выделен из проб донных отложений, отобранных в горячем источнике Гарга, расположенном в Баргузинской долине Прибайкалья. Температура воды источника на изливе достигала 75 °С на момент отбора проб. Начальное выделение штаммов из природного материала проводили на агаризованной среде Лурия-Бертани (LB). Для этого на чашки высевали по 50 мкл суспензии, культивирование проводили при температурах 60–70 °С в течение 1–3 суток. Культуры, полученные из природных образцов, очищали от сопутствующих организмов путем многократного пересева на агаризованной среде LB.

Исследование морфологии штамма проводили с использованием световых и люминесцентных микроскопов фирмы «Karl Zeiss» ЦКП микроскопического анализа биологических объектов СО РАН. Препараты готовили стандартными методами (Нетрусова, 2005).

Биохимическую характеристику штамма проводили с использованием тест-системы Enterotest-24 (MicroTest, Lachema) в соответствии с инструкцией производителя. Штамм был проверен на способность к образованию сероводорода, уреазную активность и использование следующих субстратов: лизина, орнитина, аргинина, цитрата, малоната, инозитола, адонитола, целлобиозы, сахарозы, трегалозы, маннитола, эскулина, сорбитола, рамнозы, мелибиозы, раффинозы, дульцита, глюкозы.

Препараты ДНК для полногеномного секвенирования были получены с использованием набора **genjet DNA purification kit (fermentas)** в соответствии с инструкцией производителя.

Секвенирование геномной ДНК проводили при помощи прибора MiSeq фирмы «Illumina» с использованием набора реагентов Miseq reagent kit v.2 в лаборатории эволюционной геномики факультета биоинженерии и биоинформатики Московского государственного университета им. М.В. Ломоносова.

De novo ассемблирование коротких последовательностей в контиги проводилось с использованием пакета программ CLC Genomics workbench v.6.0.4, использующего алгоритм, основанный на графах де Брюйна (de Bruijn graphs). В последующем анализе использовали контиги длиной не менее 1000 нуклеотидов.

Сравнение контигов с базой нуклеотидных последовательностей NT проводили с помощью BLASTN. Поиск открытых рамок считывания в контигах и соответствующих им потенциальных белков проводили с помощью BLASTX, который осуществляет трансляцию нуклеотидных последовательностей в аминокислотные и их сравнение с базой белковых последовательностей NR. С помощью BLASTX также проводили сравнение потенциальных белков, кодируемых рамками считывания, с известными белками *Geobacillus*. Для этого была сформирована выборка последовательностей белков *Geobacillus*, на основе которой была создана индексированная база данных в формате BLAST. Сравнение

аминокислотных последовательностей с базой данных последовательностей белков NR проводили с помощью программы BLASTP. В работе были использованы локальные версии программ пакета BLAST (<ftp://ftp.ncbi.nlm.nih.gov/blast/>).

Точную филогенетическую идентификацию полученного штамма проводили при помощи анализа 16S рРНК, для этого использовали построение филогенетического древа при помощи метода минимальной эволюции, реализованного в пакете программ MEGA 5.

Поиск типовых штаммов видов рода *Geobacillus* и соответствующих номеров последовательностей генов 16S рРНК проводили в базе данных StrainInfo (www.straininfo.net). Последовательности генов 16S рРНК брали из базы данных GenBank (www.ncbi.nlm.nih.gov/nucleotide).

Основные расчеты проводились на вычислительном кластере ЦКП «Биоинформатика» ИЦиГ СО РАН.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

В 2007 г. Институтом цитологии и генетики СО РАН была проведена экспедиция на горячий источник Гаргинский, расположенный в долине р. Баргузин в Прибайкалье. В ходе экспедиции были отобраны образцы воды и донных отложений источника. В ходе выделения были получены штаммы, определенные как относящиеся к роду *Geobacillus*.

Исследуемый штамм образует округлые колонии кремового цвета. Края колоний волнистые или слегка волнистые. Профиль колоний слегка выпуклый. Размеры колоний варьируют от 3 до 5 мм. Клетки штамма представлены палочками, размеры которых составили 2–3 × 7–10 мкм (табл. 1). У исследуемого штамма выявлено спорообразование. Эндоспоры овальные или сферические 1,5 мкм (рис. 1).

Для исследуемого штамма проведено изучение биохимических характеристик. Штамм обладал способностью к росту на средах с комплексными субстратами. Спектр соединений, утилизируемых штаммом в аэробных и анаэробных условиях, представлен в табл. 2.

Установлено, что исследуемый штамм не образует сероводород и не использует большинство предложенных субстратов, однако использует глюкозу и эскулин. Штамм дает положительную реакцию на уреазу и β-галактозидазу.

Для точной видовой идентификации выделенного штамма был проведен филогенетический анализ. По предварительным данным анализа полученной в результате секвенирования последовательности гена 16S рРНК штамм был отнесен к роду *Geobacillus* (табл. 3).

Для более точной его идентификации было проведено филогенетическое сравнение с последовательностями 16S рРНК типовых штаммов видов рода *Geobacillus* (рис. 2).

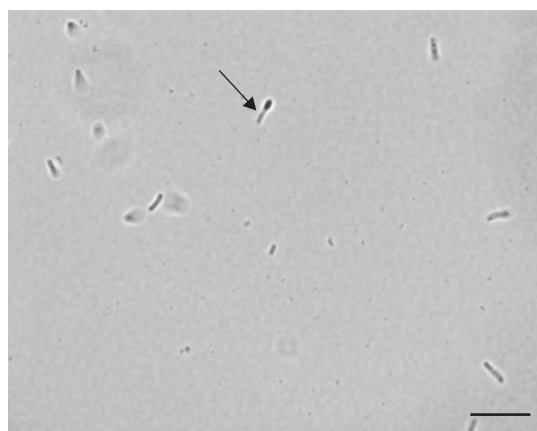


Рис. 1. Микрофотография клеток штамма *G. stearothermophilus* 22.

Клетка со спорой показана стрелкой. Масштабный отрезок – 10 мкм.

Таблица 1

Морфологическая характеристика штамма

Штамм	Морфология колоний					Морфология клеток	
	Форма	Размер, мм	Цвет	Профиль	Край	Морфотип клеток	Размеры клеток, мкм
22	круглая	3–5	кремовый	слегка выпуклый	слегка волнистый	палочки	2–3 × 7–10 споровые, d = 1,5

Таблица 2
Биохимическая характеристика штамма *G. stearothermophilus* 22

Субстрат/ Активность	Штамм 22	Субстрат/ Активность	Штамм 22
Сероводород	-	Сахароза	-
Лизин	-	Трегалоза	-
Орнитин	-	Маннитол	-
Уреаза	+	Эскулин	+
Аргинин	-	Сорбитол	-
Цитрат Симмонса	-	Рамноза	-
Малонат	-	Мелибиоза	-
β-галактозидаза	+	Раффиноза	-
Инозитол	-	Дульцит	-
Адонитол	-	Глюкоза	+
Целлобиоза	-	Индол	-
Сахароза	-	Фенилаланин	-
Целлобиоза	-	Ацетоин	-

Сравнение с последовательностями 16S рРНК типовых штаммов видов рода *Geobacillus* показало, что последовательность штамма *Geobacillus* 22 наиболее близка к последовательности типового штамма *G. stearothermophilus*. Таким образом, мы относим штамм 22 к виду *G. stearothermophilus*.

БИОИНФОРМАТИЧЕСКИЙ АНАЛИЗ

В настоящее время в базах данных отсутствует расшифрованный полный геном *G. stearothermophilus*. Сравнение контигов штамма 22 с базой данных нуклеотидных последовательностей Европейской молекулярно-биологической лаборатории EMBL (<http://www.ebi.ac.uk/embl/>) показало высокую степень сходства анализируемого штамма с геномом *G. kaustophilus* Y412MC52 (идентичность составила 93,6 %). В связи с этим для идентификации потенциальных белков в штамме 22 использовали

Таблица 3
Уровень сходства между последовательностью гена 16S рРНК штамма *G. stearothermophilus* 22 и последовательностями гена 16S рРНК типовых штаммов

Последовательность типового штамма	Количество идентичных нуклеотидов при попарном выравнивании (%)
<i>G. stearothermophilus</i> (AB021196)	1349/1352 (99)
<i>G. subterraneus</i> (AF276306)	1341/1352 (99)
<i>G. subterraneus aromaticivorans</i> (HE613733)	1336/1352 (99)
<i>G. gargensis</i> (FR749979)	1336/1352 (99)
<i>G. lituanicus</i> (AY044055)	1333/1352 (99)
<i>G. uzenensis</i> (AF276304)	1332/1352 (99)
<i>G. kaustophilus</i> (X60618)	1326/1352 (98)

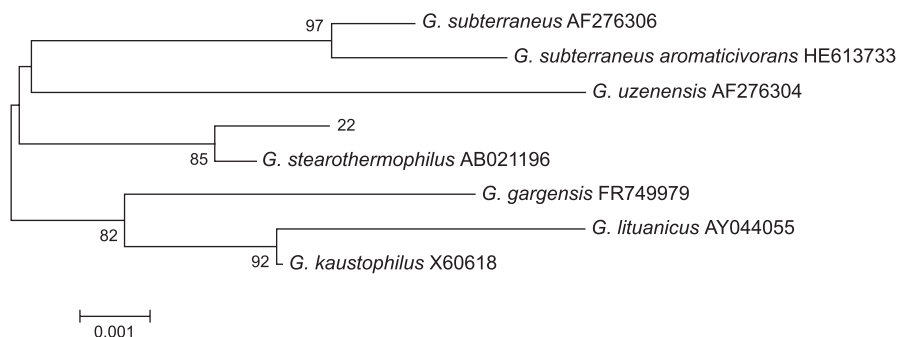


Рис. 2. Филогенетическое дерево последовательностей 16S рРНК штамма 22 и типовых штаммов видов рода *Geobacillus*, построенное методом минимальной эволюции в программе MEGA v.5.0.

Цифры возле ветвей обозначают бутстрепную поддержку.

аминокислотные последовательности охарактеризованных белков *G. kaustophilus* Y412MC52 (<http://biocyc.org/GSP550542/organism-summary?object=GSP550542>). Удалось охарактеризовать 3186 открытых рамок считывания из штамма 22, которые имели высокое сходство с белками *Geobacillus* sp. Y412MC52 (всего для него известно 3639 белков).

АНАЛИЗ МЕТАБОЛИЧЕСКИХ ПУТЕЙ ШТАММА 22

Большое внимание привлекает к себе проблема получения биотоплива, в частности биоэтанола, из растительного сырья. Перспективным сырьем для выработки биоэтанола является лигноцеллюлозная биомасса, которая может быть

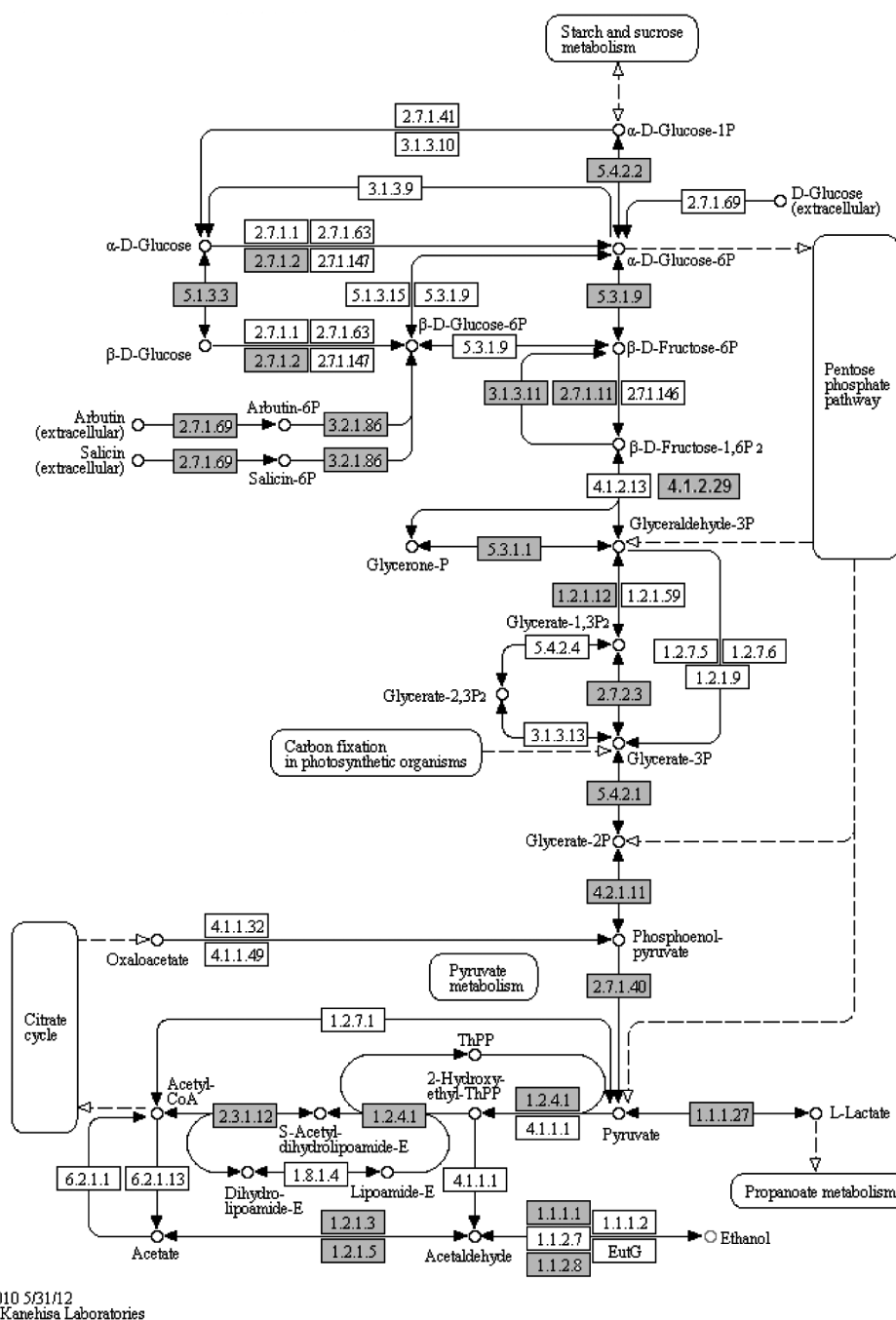


Рис. 3. Гликолитический метаболический путь, описанный в базе данных KEGG (Kanehisa, Goto, 2000; Kanehisa *et al.*, 2012).

Серыми прямоугольниками отмечены ферменты, обнаруженные у штамма *G. stearothermophilus* 22.

утилизирована различными микроорганизмами (Tanimura *et al.*, 2012; Zhu *et al.*, 2013). Однако использование природных штаммов микроорганизмов для наработки биоэтанола в промышленном масштабе затруднено в связи с низким уровнем переработки целлюлозосодержащего сырья, низким выходом этанола и образованием большого количества побочных продуктов. Наиболее перспективным путем считается гидролиз лигноцеллюлозной биомассы до сахаров с последующей ферментацией микроорганизмами. Для получения продуцента, эффективно дающего продукт метаболизма, необходимо, чтобы используемый в работе штамм получал только один, целевой, продукт в результате катаболизма. В большинстве случаев для этого необходимо проведение модификации метаболизма природных продуцентов. Для чего необходима информация об имеющихся в клетке путях катаболизма, о последовательностях ферментов, закодированных в геноме. Таким образом, реконструкция и анализ метаболических путей могут быть основой для проведения генных модификаций, обеспечивающих получение мутантных штаммов микроорганизмов, способных с высокой эффективностью перерабатывать гидролизаты лигноцеллюлозной биомассы до этанола.

Ключевыми для наработки биоэтанола являются реакции гликолитического метаболического пути (рис. 3). В основном это цепочка превращений D-глюкозы-1-фосфат в D-глюкозу-6-фосфат, затем в бета-D-фруктозу-6-фосфат и бета-D-фруктозу-1,6-бисфосфат, далее в глицеральдегид-3-фосфат, глицерат 1,3-дифосфат, 3-фосфоглицерат и 2-фосфоглицерат, фосфоенолпируват, пируват, ацетил-CoA и затем в этанол (рис. 3). Практически все основные ферменты, ответственные за образование этанола из растительного сырья, были выявлены у вновь секвенированного штамма *G. stearothermophilus* 22 (рис. 3) на основе полученных последовательностей.

Кроме того, в геноме бактерии обнаружены фермент гидролиза целлюлозы, эндо-1,4-бета глюконаза и ферменты, участвующие в гидролизе гемицеллюлозы, – эндо-1,4-бета ксиланаза, бета ксилозидаза и альфа глюкононидаза. Присутствие в геноме этих ферментов говорит о возможности использования данного штамма для гидролиза гемицеллюлозы, которая могла

попадать в его среду обитания вместе с опадающими листьями, богатыми ксиланами.

Таким образом, у вновь секвенированного штамма *G. stearothermophilus* 22, обитающего в природном термальном источнике Гарга, расположенном в Баргузинской долине Прибайкалья, были выявлены все ключевые ферменты, ответственные за синтез этанола из растительного сырья, что открывает возможность для использования в промышленности штамма *G. stearothermophilus* 22 для производства биотоплива. Штамм обладает ограниченным набором ферментов карбогидраз, что может положительно сказаться при ферментации лигноцеллюлозных гидролизатов.

Работа выполнена при финансовой поддержке Министерства образования и науки Российской Федерации (ГК 14.512.11.0072 от 19.04.2013 г.) в рамках ФЦП «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007–2013 гг.».

ЛИТЕРАТУРА

- Бонч-Осмоловская Е.А., Мирошниченко М.Л., Соколова Т.Г., Слободкин А.И. Термофильные микробные сообщества: новые физиологические группы, новые местообитания // Тр. Ин-та микробиологии им. С.Н. Виноградского. М.: Наука, 2004. Вып. 12.
- Нетрусова А.И. Практикум по микробиологии. М.: Академия, 2005. 608 с.
- Brock T.D. Thermophilic microorganisms and life at high temperatures. N.Y.: Springer-Verlag, 1978. 465 p.
- Cripps R.E., Eley K., Leak D.J. *et al.* Metabolic engineering of *Geobacillus thermoglucosidarius* for high yield ethanol production // Metab. Eng. 2009. V. 11. P. 398–408.
- Kanehisa M., Goto S. KEGG: kyoto encyclopedia of genes and genomes // Nucl. Acids Res. 2000. V. 28. P. 27–30.
- Kanehisa M., Goto S., Sato Y. *et al.* KEGG for integration and interpretation of large-scale molecular datasets // Nucl. Acids Res. 2012. V. 40. P. D109–D114.
- Suzuki H., Yoshida K. Genetic transformation of *Geobacillus kaustophilus* HTA426 by conjugative transfer of host-mimicking plasmids // J. Microbiol. Biotechnol. 2012. V. 22. P. 1279–1287.
- Tanimura A., Nakamura T., Watanabe I. *et al.* Isolation of a novel strain of *Candida shehatae* for ethanol production at elevated temperature // Springerplus. 2012. V. 4. P. 1–27.
- Taylor M.P., Esteban C.D., Leak D.J. Development of a versatile shuttle vector for gene expression in *Geobacillus* ssp. // Plasmid 60. 2008. P. 45–52.
- Zhu X., Cui J., Feng Y. *et al.* Metabolic adaptation of ethanol-tolerant *Clostridium thermocellum* // PLoS ONE. 2013. V. 8. P. E70631.

**BIOINFORMATIC ANALYSIS OF THE GENOME
OF THE *GEOBACILLUS STEAROTHERMOPHILUS* 22 STRAIN
ISOLATED FROM THE GARGA HOT SPRING, BAIKAL REGION**

**A.S. Rozanov¹, T.V. Ivanisenko¹, A.V. Bryanskaya¹, S.V. Shekhovtsov¹,
M.D. Logacheva², O.V. Saik¹, T.K. Malup¹, P.S. Demenkov¹,
T.N. Goryachkovskaya¹, V.A. Ivanisenko¹, S.E. Peltek¹**

¹ Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia,
e-mail: peltek@bionet.nsc.ru;

² Lomonosov Moscow State University, Moscow, Russia

Summary

A new strain, *Geobacillus stearothermophilus* 22 was isolated from the Garga hot spring in the Bargusin Valley, Baikal Region, Russia. The morphology and biochemistry of the strain were analyzed, and the genome-wide sequencing was conducted. The sequence was subjected to bioinformatic analysis. Nucleotide sequences (contigs) of the strain were found to be similar to the genome of the thermophilic strain *G. kaustophilus* Y412MC52. The proteome of the new strain was analyzed. Fragments associated with hemicellulases (endoxylanase, beta xylosidase, and arabinofuranosidase) and the endoxylanase enzyme were detected.

Key words: *Geobacillus*, thermophile, genome-wide sequencing, bioinformatic analysis.