

УДК 579.258+575.852

СРАВНИТЕЛЬНАЯ ГЕНОМИКА УТИЛИЗАЦИИ УГЛЕВОДОВ У БАКТЕРИЙ СЕМЕЙСТВА *SPHAEROCHAETACEAE*: ЭВОЛЮЦИОННОЕ ПРОИСХОЖДЕНИЕ ГЕНОВ ГАЛАКТУРОНИДАЗЫ И НЕНАСЫЩЕННОЙ РАМНОГАЛАКТУРОНИЛ-ГИДРОЛАЗЫ¹

© 2024 г. О. Ю. Трошина^{а, *}, Д. Г. Наумов^б, В. И. Речкина^а, В. А. Щербакова^а

^аФИЦ “Пушкинский научный центр биологических исследований РАН”,

Институт биохимии и физиологии микроорганизмов им. Г.К. Скрыбина РАН, Пушкино, 142290, Россия

^бИнститут микробиологии им. С.Н. Виноградского, ФИЦ Биотехнологии РАН, Москва, 119071, Россия

*e-mail: oltro676@yandex.ru

Поступила в редакцию 27.02.2024 г.

После доработки 08.04.2024 г.

Принята к публикации 09.04.2024 г.

Проведен сравнительный анализ белков деградации углеводов, закодированных в доступных в настоящее время геномных последовательностях бактерий семейства *Sphaerochaetaceae*, а именно *Sphaerochaeta associata* GLS2^T, *S. globosa* Buddy^T, *S. pleomorpha* Grapes^T, *S. halotolerans* 4-11^T, *S. halotolerans* 585, *Sphaerochaeta* sp. S2, *Sphaerochaeta* sp. PS и *Parasphaerochaeta coccoides* SPN1^T. Геномы сферохет кодируют средний по размеру и по разнообразию набор белков, потенциально участвующих в деградации разных классов углеводов, в основном олигосахаридов. Все исследованные геномы кодируют гликозил-гидролазы семейств GH1, GH2, GH3, GH4, GH13, GH20, GH28, GH36, GH43, GH57, GH63, GH77, GH105 и карбогидратэстеразы семейств CE8 и CE9. У всех исследованных сферохет, за исключением *P. coccoides* SPN1^T, в геномах закодировано много белков семейства GH31. У исследованных представителей *Sphaerochaetaceae* нет генов эндо-β-ацетилмурамидазы (лизоцима) семейства GH23, вовлеченной в процесс оборота пептидогликана. Однако геномы *S. associata*, *S. globosa*, *Sphaerochaeta* sp. PS и *S. pleomorpha* содержат ген экзо-β-ацетилмурамидазы (семейство GH171). Значительная часть генов, кодирующих ферменты деградации углеводов, имеет наиболее близкие гомологи среди представителей филумов *Bacillota*, *Bacteroidota*, *Pseudomonadota*. В геномах исследованных бактерий закодированы белки, которые потенциально могут участвовать в деградации пектина. Исследованы способность представителей *Sphaerochaetaceae* использовать для роста пектин, а также эволюционное происхождение генов, кодирующих потенциальные α-галактуронидазу (семейство GH4) и ненасыщенную глюкуронил/рамногалактуронил-гидролазу (семейство GH105), которые участвуют в деградации компонентов пектина.

Ключевые слова: *Sphaerochaeta*, горизонтальный перенос генов, гликозил-гидролаза, полисахаридлиаза, CAZy, GH4, GH105, α-галактуронидаза, ненасыщенная глюкуронил/рамногалактуронил-гидролаза

DOI: 10.31857/S0026365624050034

Сравнительные исследования геномов микроорганизмов показали, что они содержат гены с разной эволюционной историей. Кроме высоко консервативных генов, так называемых “ядерных” или универсальных генов, которые наследуются главным образом вертикально, значительная доля генов, кодирующих периферические метаболические функции и транспортные белки, подвергается масштабным перемещениям как между близкими видами, так и между бактериями из далеких таксонов. Приобретение устойчивости

к антибиотикам и токсинам, генов патогенности, способности использовать новые субстраты и ксенобиотики являются следствием горизонтального переноса генов. Признано, что он играет ключевую роль в эволюции адаптации прокариот к изменяющимся условиям среды (Ochman et al., 2000; Arnold et al., 2022). Ряд исследований указывает на важную роль самой окружающей среды в эволюции геномов бактерий и в их адаптации к специфической экологической нише (Саго-Quintero et al., 2011). Так, сравнительная геномика

¹Дополнительные материалы к статье размещены по ссылке DOI: 10.31857/S0026365624050034

штаммов молочнокислых бактерий *Lactobacillus helveticus* DPC4571 и *L. acidophilus* NCFM, имеющих 98.4% идентичности на уровне генов 16S рРНК, выявила наличие у них специфичных генов или баркодов, типичных для определенной экониши этих бактерий, молока или кишечника животных соответственно (O'Sullivan et al., 2009). Значительная доля генов, полученных в результате горизонтального переноса, отмечена для геномов кокковидных неподвижных спирохет *Sphaerochaeta globosa* и *S. pleomorpha* из семейства *Sphaerochaetaceae*. “Химерные” геномы этих двух бактерий содержат более 10% универсальных генов (генов ядра) и более 50% генов, кодирующих вторичные метаболические функции и транспортные белки, полученные, по-видимому, от грамположительных бактерий *Bacillota* (Caro-Quintero et al., 2012). Большинство генов приобретенных путем горизонтального переноса у исследованных сферохет относятся к категориям транспорта и метаболизма углеводов и аминокислот (Caro-Quintero et al., 2012).

В настоящее время в семействе *Sphaerochaetaceae* известны три рода, включающих восемь видов, а именно *Sphaerochaeta associata* GLS2^T, *S. globosa* Buddy^T, *S. halotolerans* 4-11^T, *S. pleomorpha* Grapes^T, *Parasphaerochaeta coccoides* SPN1^T, *Pleomorphochaeta caudata* SEBR 4223^T, *P. naphthae* SEBR 4209^T и *P. multiformis* MO-SPC2^T (Nazina et al., 2022). К этому семейству принадлежат также некультивируемые виды родов “*Candidatus Aphodenecus*”, “*Ca. Ornithospirochaeta*”, “*Ca. Physcosoma*”, описанные на основе метагеномных последовательностей, полученных из кишечного тракта птиц и млекопитающих (Gilroy et al., 2021, 2022; Gharechahi et al., 2022). Предлагается выделить семейство *Sphaerochaetaceae* в отдельный порядок *Sphaerochaetales* в классе *Spirochaetia* (Chuvochina et al., 2023). Последовательности генов сферохет широко представлены в разнообразных анаэробных местообитаниях: морских и пресноводных осадках, сточных водах, в нефтяных пластах, в биореакторах по переработке отходов, получения биогаза, в кишечном тракте насекомых, птиц и млекопитающих (Morrison et al., 2020; Tang et al., 2020; Gilroy et al., 2021, 2022; Gharechahi et al., 2022; Nazina et al., 2022). Доступность полногеномных последовательностей близкородственных представителей *Sphaerochaetaceae*, выделенных из разнообразных экологических ниш, позволяет понять, существует ли адаптация этих видов к специфической среде обитания на геномном уровне, а также изучить происхождение некоторых генов.

Все известные представители *Sphaerochaetaceae*, также как и таксономически близкие им бактерии из семейства *Spirochaetaceae*, являются хемоорганогетеротрофами и используют углеводы в качестве основного источника углерода и энергии. Это

предопределило применение нами в работе международной базы данных ферментов синтеза и утилизации углеводов (Carbohydrate-Active Enzymes, CAZy, <http://www.cazy.org/>), а точнее ее разделов, посвященных трем группам ферментов: гликозил-гидролазам (GH), полисахаридлиазам (PL) и карбогидратэстеразам (CE) (Drula et al., 2022). Ферменты синтеза и деградации углеводов, классифицированные в базе данных CAZy, названы далее термином CAZy-мы. Для каждой из этих групп ферментов классификация CAZy выделяет семейства на основании гомологии аминокислотных последовательностей их каталитических доменов (Drula et al., 2022). Так, в случае гликозил-гидролаз различают почти две сотни семейств (GH1–GH189), более трети из которых объединены в 20 кланов (GH-A–GH-S-T). В пределах части семейств принято выделять подсемейства (например, 49 подсемейств в семействе GH13: GH13_1–GH13_49 или 39 подсемейств в семействе GH43). В предложенной нами ранее иерархической классификации гликозил-гидролаз (Наумов, 2011) почти все разнообразие каталитических доменов этих ферментов сведено к шести основным фолдам — крупным группам белков, имеющих однотипную пространственную структуру и предполагаемую общность эволюционного происхождения. В частности, гликозил-гидролазы с каталитическим доменом в виде (α/α)₆-бочонка представлены в семи кланах (GH-G, GH-L, GH-M, GH-O, GH-P, GH-Q и GH-S), а также в целом ряде семейств, не относящихся к кланам, в т.ч. в семействе GH105. Это семейство объединяет ферменты, отщепляющие от своих субстратов ненасыщенные глюкоуронильные или галактуронильные остатки. Гены белков этого семейства представлены в геномах бактерий и грибов. Семейство GH4 является редким примером семейства, не относящегося ни к одному из шести основных фолдов, а характеризуется наличием у него так называемой укладки Россманна. Ферменты этого и родственного ему семейства GH109 характеризуются необычным для гликозил-гидролаз молекулярным механизмом с использованием НАД⁺ в качестве кофактора. Для ферментов семейства GH4 показано большое разнообразие катализируемых ими активностей (Hall et al., 2009). Проведенный нами ранее филогенетический анализ позволил выделить в нем четыре монофилетические группы белков, характеризующихся различающимися ферментативными активностями (Березина и соавт., 2003). Гены белков этого семейства представлены в геномах бактерий и архей. В 2013 году была установлена еще одна ферментативная активность для ряда белков семейства GH4, а именно активность α-галактуронидазы (Thompson et al., 2013).

Анализ в базе CAZy профилей ферментов деградации и синтеза углеводов (CAZy-мов)

у некоторых известных представителей *Spirochaetia* показывает, что они в значительной степени коррелируют с типом экологической ниши обитания микроорганизмов. *Treponema rectale*, *T. bryantii*, изолированные из кишечного тракта быка, *Thiospirochaeta perfilievi*, *Spirochaeta thermophila*, *Gracilinema caldarium*, *Salinispira pacifica*, выделенные из микробных матов, имели разнообразный и широкий набор CAZy-белков в отличие от родственных им спирохет *Entomospira culicis* и *E. entomophilus*, изолированных из комаров, а также многих патогенных трепонем и боррелий.

Цель настоящей работы заключалась в сравнительном исследовании геномных профилей ферментов деградации углеводов бактерий *Sphaerochaetaceae*, изучении способности представителей этого семейства использовать пектин в качестве источника углерода и анализ эволюционного происхождения генов потенциальных α -галактуронидазы и ненасыщенной глюкоуронил/рамногалактуронил-гидролазы.

МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЯ

В работе использовали полные геномные последовательности *S. associata* GLS2^T (GCF_022869165.1), *S. globosa* Buddy^T (GCF_000190435.1), *S. pleomorpha* Grapes^T (GCF_000236685.1), и *P. coccoides* SPN1^T (GCF_000208385.1) (Abt et al., 2012; Caro-Quintero et al., 2012; Troshina et al., 2024), а также геномные данные хорошего качества для *S. halotolerans* 4-11^T (GCA_003429665.1) (Grouzdev et al., 2018), *S. halotolerans* 585 (GCF_009823595.1) и изолятов *Sphaerochaeta* sp. P5 (GCA_032144095.1), *Sphaerochaeta* sp. S2 (GCA_016406725.1).

Идентификацию CAZy-мов, кодируемых в геномах исследуемых бактерий, проводили с помощью ресурса dbCAN3 (<http://bcbl.unl.edu/dbCAN2/>), используя три алгоритма — HMMER:dbCAN, DIAMOND:CAZy и HMMER:dbCAN-sub (Zheng et al., 2023). Только белки, найденные хотя бы двумя из алгоритмов, рассматривали как CAZy-мы. Для сравнения использовали также данные CAZy для геномов *S. associata* GLS2^T, *S. globosa* Buddy^T, *S. pleomorpha* Grapes^T, *P. coccoides* SPN1^T (<http://www.cazy.org/>) (Drula et al., 2022). Предсказание сигнальных пептидов проводили с dbCAN3 и Signal-BLAST (Frank, Sippl, 2008). Номера всех идентифицированных ферментов деградации углеводов исследованных геномов приведены в таблице S1 в приложении.

Поиск гомологов CAZy-мов, кодируемых в геномах *Sphaerochaetaceae*, проводили с помощью BLASTP (<https://blast.ncbi.nlm.nih.gov/>) в базе данных NCBI (раздел “non-redundant protein sequences”). При составлении таблицы

по таксономическому положению ближайших гомологов выбирали только гомологи, принадлежащие организмам с четким таксономическим положением, с покрытием не менее 80%. Если ближайшие гомологи принадлежали одному виду, то отбирали только первую ближайшую последовательность. Для построения филогенетических деревьев отбирали гомологи, имеющие максимальное покрытие исследуемого белка и с четким таксономическим положением, а также ряд метагеномных последовательностей организмов семейства *Sphaerochaetaceae*. Для выравнивания последовательностей, поиска моделей и построения филогенетических деревьев использовали пакет программ MEGA11 (Tamura et al., 2021). Деревья построены методом наибольшего правдоподобия (ML) по модели Le_Gascuel_2008 с дискретным гамма-распределением для моделирования различий в скорости эволюции между участками (5 категорий (+G) и модель изменения скорости позволила некоторым сайтам быть эволюционно неизменными [+I]). Статистическую надежность узлов филогенетических деревьев оценивали с помощью бутстреп-анализа, используя 500 реплик. Для презентации филогенетических деревьев использовали программу TreeViewer (<https://treeviewer.org>) (Bianchini, Sánchez-Baracaldo, 2024). Для идентификации гомологичных генных кластеров, кодирующих белки семейств GH88, PL12 и PL33, использовали сервис CAGECAT (<https://cagecat.bioinformatics.nl/#>) (Gilchrist et al., 2021). Анализ транспортных систем и ферментов метаболизма уроновых кислот проводили с помощью ресурсов KEGG и BlastKOALA (Kanehisa et al., 2016).

Культуры *S. associata* GLS2^T, *S. globosa* Buddy^T, *S. pleomorpha* Grapes^T и *P. coccoides* SPN1^T выращивали анаэробно в атмосфере N₂ в пробирках Хангейта на среде Пфеннинга (г на 1 л дистиллированной воды): NH₄Cl — 0.33, KH₂PO₄ — 0.33, KCl — 0.33, MgCl₂ · 6H₂O — 0.5, NaCl — 1.0, CaCl₂ · 6H₂O — 0.15, дрожжевой экстракт — 0.5, казаминовые кислоты — 0.5, витамины DSMZ 503 — 1 мл, микроэлементы — 1 мл (Кевбрин, Заварзин, 1992), L-цистеин гидрохлорид — 0.4, пектин (полигалактуронозная кислота, CDH) — 3.0. Для *P. coccoides* SPN1 количество дрожжевого экстракта и казаминовых кислот в среде было увеличено до 1.0 г на 1 л. Перед посевом в среду дополнительно вносили (г/л): Na₂S · 9H₂O — 0.25, NaHCO₃ 1.0–2.0. Культивирование проводили при 29°C. Рост оценивали после четырех пересевов на среду с пектином путем измерений оптической плотности при длине волны 600 нм на спектрофотометре модели Cary 300 (“Agilent Technologies”, Австралия) и путем прямых подсчетов клеток с помощью фазово-контрастной микроскопии, используя микроскоп AxiostarPLUS (“Carl Zeiss”, Германия).

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Проведенный анализ показал, что данные, полученные в результате аннотации белков, кодируемых в геномах *S. associata*, *S. globosa*, *S. pleomorpha* и *P. coccoides* с помощью dbCAN3, и данные, приведенные для вышеуказанных геномов в базе CAZy, совпали, за исключением нескольких карбоксиэстераз из семейств CE1, CE4, CE14. Все исследованные представители *Sphaerochaetaceae* имеют средний по размеру и в основном сходный набор семейств CAZy с незначительными различиями в числе генов каждого семейства и в профилях CAZy семейств (таблица). Все исследованные геномы кодируют белки гликозил-гидролаз семейств GH1, GH2, GH3, GH4, GH13, GH20, GH28, GH36, GH43, GH57, GH63, GH77, GH105 и карбогидратэстераз CE8 и CE9. У всех сферохет, за исключением *P. coccoides* SPN1^T, в геномах закодировано

много белков семейства GH31. Геномы *S. associata* GLS2^T, *S. globosa* Buddy^T и изолята *Sphaerochaeta* sp. S2 кодируют наибольшее число белков CAZy семейств и число CAZy-мов (таблица).

Наиболее представленные гены белков семейств GH2, GH3, GH4, GH31 кодируют потенциальные β-галактозидазы, β-глюкозидазы, α-глюкозидазы, α-галактозидазы, α-ксилозидазы. У всех исследованных сферохет присутствуют гены, кодирующие белки, вероятно, обеспечивающие деградацию соединений крахмала. Многие из этих белков относятся к семейству GH13. Согласно данным CAZy, белки семейства GH13 закодированы также в геномах многих представителей семейства *Spirochaetaceae*, например, в геноме *Spirochaeta africana* содержится 31 ген, у *Salinispira pacifica* их 22 (<http://www.cazy.org/>).

Анализ CAZy-мов сферохет на предмет таксономического положения наиболее близких гомологов программой BLASTP показал, что многие

Таблица. Численность закодированных белков из семейств CAZy для геномов *S. associata* GLS2^T (1), *S. globosa* Buddy^T (2), *S. halotolerans* 4-11^T (3), *S. halotolerans* 585 (4), *S. pleomorpha* Grapes^T (5), *Sphaerochaeta* sp. PS (6), *Sphaerochaeta* sp. S2 (7), *P. coccoides* SPN1^T (8)

CAZy семейства	Геномы							
	1	2	3	4	5	6	7	8
GH1	1	1	1	3	1	1	1	2
GH2	5	5	2	4	3	3	5	5
GH3	4	4	3	3	4	4	4	1
GH4	5	3	3	2	1	3	3	1
GH5	0	0	0	0	0	0	1	0
GH9	0	0	0	0	1	0	0	0
GH13	4	4	6	6	6	4	4	4
GH15	0	0	0	0	1	0	0	0
GH20	2	2	2	2	2	1	2	1
GH28	1	2	1	1	1	1	2	1
GH29	1	1	0	0	0	1	0	1
GH30	0	0	1	0	2	0	1	0
GH31	5	5	3	4	6	4	4**	0
GH36	1	1	1	1	1	1	1	2
GH38	1	0	1	3	0	0	0	1
GH42	0	0	0	1	0	0	1	0
GH43	4	2	2	2	1	1	2	1
GH50	0	0	0	0	1	0	0	0
GH51	1	1	1	1	0	1	1	1
GH52	0	0	0	0	0	0	0	1
GH57	2	2	2	2	2	2	3	3
GH57***	1	1	1	1	1	1	1	1
GH59	0	0	0	0	0	0	0	1
GH63	1	1	1	1	1	1	1	1
GH65	0	0	2	2	2	1	2	1
GH66	0	0	0	0	1	0	0	0
GH67	0	0	0	0	0	0	0	1

Окончание таблицы

CAZy семейства	Геномы							
	1	2	3	4	5	6	7	8
GH77	2	2	2	2	2	2	2	2
GH78	1	1	1	1	0	0	1	0
GH88	2	1	0	0	0	1	1	1
GH94	1	1	0	0	0	0	0	0
GH97	0	0	0	0	0	0	0	1
GH105	4	3	3	3	1	2	2	3
GH106	1	0	0	0	0	1	0	0
GH109	0	0	0	0	0	1	0	0
GH112	2	1	0	0	0	0	1	0
GH125	0	0	1	1	0	0	1	0
GH127	1	2	0	1	0	2	1	0
GH130	0	0	1	1	0	0	0	0
GH151	0	0	0	0	1	1	0	0
GH171	1	1	0	0	1	1	0	0
GH172	0	0	0	0	0	0	0	1
GH179	0	0	0	0	0	0	0	1
PL12	3	1	0	0	1	1	1	2
PL26	0	0	0	0	0	0	0	1
PL33	1	1	0	0	0	1	1	1
PLnc	1	1	0	0	1*	1	1	0
CE1	1*	1*	0	0	1*	1	0	0
CE4	1*	1*	1	1	1*	0	0	0
CE8	1	1	1	1	1	1	1	1
CE9	3	3	1	1	3	3	2	1
CE12	1	1	1	1	0	0	1	1
CE14	1*	1*	1	0	1	1	1	0
CE15	0	0	0	0	1	0	0	0
CE20	0	1	0	0	0	0	0	0
Число генов GH + PL + CE	67	59	46	52	53	50	56	46
Число CAZy семейств	34	33	27	27	31	31	32	31

*Отмеченные белки не присутствуют в списках CAZy. Эти белки были аннотированы программой dbCAN3.
**Один из белков семейства GH31 (WP_304945009.1) представлен в виде фрагмента. Он не был идентифицирован с помощью ресурса dbCAN3, он был идентифицирован BLAST.
***Белки отнесены к семейству GH57 программой dbCAN3. В базе CAZy гомологичные белки у *S. associata* GLS2^T, *S. globosa* Buddy^T, *S. pleomorpha* Grapes^T и *P. coccoides* SPN1^T отнесены к GHnc (неклассифицированные).

гены, кодирующие белки деградации углеводов, были, по-видимому, получены в результате горизонтальных переносов из таксономически далеких таксонов *Bacillota*, *Pseudomonadota* и *Bacteroidota* (табл. S2 в приложении). Например, белки WP_244773140.1, WP_013606124.1, WP_117331179.1 и WP_198892871.1 (вероятная α-галактозидаза, семейство GH4) имеют высокий уровень сходства (около 74% идентичности) с гомологичными последовательностями из *Bacillota*. Только белки

семейств GH1, GH57, GH63, GH77 и некоторые белки семейств GH4 и GH13 имеют среди ближайших гомологов последовательности из *Spirochaetota* (табл. S2 в приложении).
У всех исследованных сферохет, в отличие от других известных представителей *Spirochaetota*, не обнаружены гены, кодирующие ферменты оборота пептидогликана, а именно пептидогликан-лиазы и мурамидазы/лизосима, которые относятся к семейству GH23. Однако *S. associata*,

S. globosa, *S. pleomorpha* и *Sphaerochaeta* sp. PS содержат гены, кодирующие белки с относительно высоким уровнем сходства (около 40% идентичности) к недавно охарактеризованному ферменту деградации пептидогликана экзо- β -N-ацетилмурамидазе (семейство GH171; P40407.1 в базе <https://www.uniprot.org/>) (Müller et al., 2021). Ближайшими гомологами этих белков, кодируемых в геномах вышеупомянутых сферохет, являются последовательности из *Bacillota* (табл. S2 в приложении). Гены белков семейства GH171 есть также у ряда представителей рода *Treponema*, а у многих представителей семейства *Leptospiraceae* (класс *Spirochaetia*) даже по 2 копии этого гена. Следует отметить, что набор предсказанных гликозил-гидролаз у *Leptospiraceae* (данные из CAZy), как патогенных, так и свободноживущих, является очень небольшим.

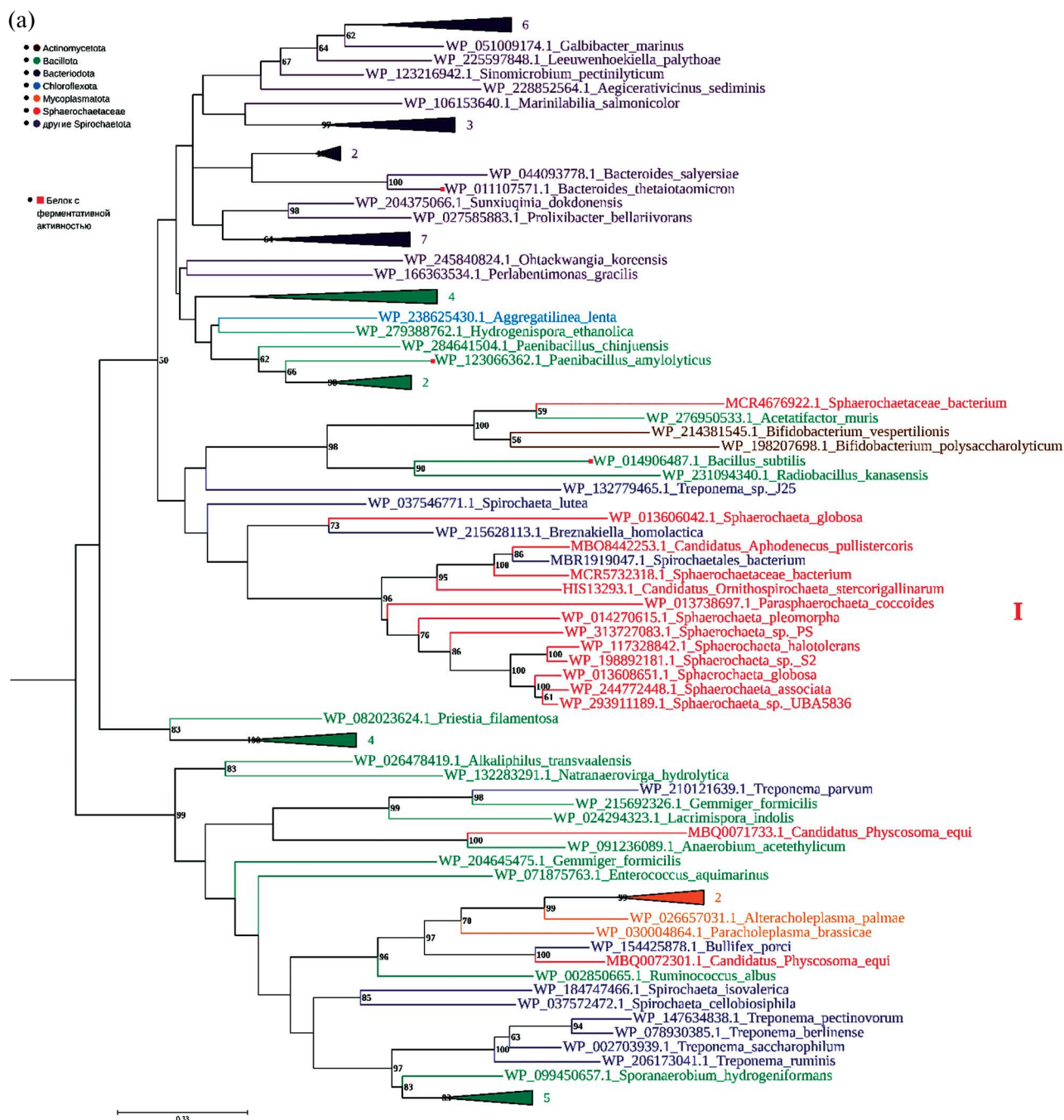
По сравнению с другими сферохетами геном *S. associata* GLS2^T имеет наибольшее число генов, кодирующих белки семейств GH43, GH88, GH105, PL12 и PL33 (таблица). Эти белки аннотированы как потенциальные ксилан-1,4- β -ксилозидазы (GH43), α -N-арабинофуранозидаза (GH43), ненасыщенная глюкуронил-гидролаза (семейство GH88), гепариназа II/III подобные белки (семейства PL12, PL33). Белки WP_244772593.1 и WP_244774031.1 из семейства GH88 у *S. associata* GLS2^T обнаруживают, соответственно, 46.1 и 45.0% идентичности в аминокислотной последовательности с биохимически охарактеризованным ферментом хондроитиндисахарид-гидролазой из *Streptococcus agalactiae* (номер Q8E372.1 в базе <https://www.uniprot.org/>). Белок WP_244774018.1 (семейство PL33) *S. associata* GLS2^T имеет 31.5% идентичности с белком AEW20097.1 из бактерии *Tannerella forsythia* (филум *Bacteroidota*), для которого была показана хондроитин-лиазная активность (К.Ф. 4.2.2.5) (Nguyen et al., 2022). Гены, кодирующие белки семейств GH88, PL12 и PL33, находятся рядом в геномах *S. associata* GLS2^T, *S. globosa* Buddy^T, *P. coccoides* SPN1^T и *Sphaerochaeta* sp. PS (рис. S1 в приложении), что может свидетельствовать об их функциональной взаимосвязи. Кластер генов вышеуказанных белков, обнаруживаемый в геномах многих сферохет, содержит также гены ABC-транспортных систем углеводов и гены белков двухкомпонентной регуляторной системы транскрипции YesM/YesN (рис. S1). Ген еще одной глюкуронил-гидролазы семейства GH88 (WP_244772593.1) *S. associata* GLS2^T также ко-локализуется с геном потенциальной гепариназы семейства PL12 (рис. S2). В этом локусе есть также гены сульфатаз и ABC-транспортных систем углеводов (белки WP_244772597.1, WP_244772598.1 и WP_244772599.1) (рис. S2 в приложении). Белки транспорта углеводов, закодированные в данном кластере генома *S. associata* GLS2^T, имеют высокую

степень сходства (более 70% идентичности) с гомологами из *Bacillota*. Потенциальная сульфатаза из *S. associata* GLS2^T (WP_244772595.1) обнаруживает 33.0% сходства с охарактеризованной сульфатазой из бактерии *Bacteroides thetaiotaomicron* VPI-5482 (Q8A2F6.1), для которой показан гидролиз сульфатированных глюкозаминогликанов хондроитина и дерматансульфатов (Ulmer et al., 2014). Следует отметить, что штамм *S. associata* GLS2^T был выделен из устойчивого консорциума с археей рода *Methanosarcina* (Troshina et al., 2015), которые, как известно, содержат в составе клеточных оболочек метанохондроитин, состоящий из N-ацетил-D-галактозамина и D-глюкуроновой (или D-галактуроновой) кислоты в соотношении 2 : 1 и со следующей структурой $[\rightarrow\beta\text{-D-GlcA-(1\rightarrow3)-}\beta\text{-D-GalNAc-(1\rightarrow4)-}\beta\text{-D-GalNAc-(1\rightarrow)n}]$ (Klingl et al., 2019). Белки семейств GH88, PL12, PL33 у исследованных сферохет имеют среди ближайших гомологов последовательности представителей таксонов *Pseudomonadota* и *Bacillota* (табл. S2 в приложении), и, по-видимому, гены этих белков были получены в результате горизонтальных переносов. Гепариназа II/III подобные белки семейства PL12 сферохет имеют низкое сходство (около 25% идентичности для белка WP_244772004.1) с биохимически охарактеризованными гепаринсульфат-лиазами. Можно предположить, что эти белки, при условии их выделения и изучения каталитических свойств, могут проявлять новую ферментативную активность.

В геномах *S. globosa*, *S. associata*, *Sphaerochaeta* sp. PS и *Sphaerochaeta* sp. S2 присутствует ген, кодирующий белок, который в настоящее время не классифицирован в CAZy. Его относят к семейству PLnc. В последовательности этого белка есть консервативная область рамногалактуронан-лиазы. Эти белки аннотированы алгоритмом HMMER в dbCAN3 как белки семейства PL11, но как неклассифицированные белки семейства PLnc алгоритмом DIAMOND. В геноме *P. coccoides* закодирована рамногалактуронан-экзолиаза семейства PL26. Этот белок, так же как и белок семейства PLnc, возможно, участвует в расщеплении олигосахаридов рамногалактуронана I, одного из компонентов пектина. Образующие в результате действия рамногалактуронан-лиаз по типу β -элиминирования олигосахариды с Δ -4,5-ненасыщенными α -галактуронопиранозными остатками на невозстанавливаемом конце в дальнейшем являются субстратами для белков семейства GH105. Охарактеризованные к настоящему времени белки этого семейства катализируют гидролитическое отщепление ненасыщенного остатка α -галактуроновой или α -глюкуроновой кислот от олигосахаридов пектина или ульвана соответственно (http://www.cazy.org/GH105_characterized.html). Впервые ферменты семейства GH105 были

выделены в 2006 году из *Bacillus subtilis* (Itoh et al., 2006). Гомологи потенциальных ненасыщенных рамногалактуронил-гидролаз (К.Ф. 3.2.1.172) и d-4,5-ненасыщенной α -галактуронидазы (К.Ф. 3.2.1.-) закодированы в геномах исследованных *Sphaerochaetaceae* (семейство GH105, таблица). В то же время белки семейства GH105 есть только у некоторых из аннотированных в базе CAZY представителей *Spirochaetota*, а именно только у некоторых видов родов *Treponema* и *Spirochaeta*. Напротив, у исследованных *Sphaerochaetaceae*

этот ген присутствует в количестве от 1 до четырех копий (таблица). То, что эти гены сохраняются в геномах сферохет и присутствуют в нескольких копиях, свидетельствует о возможном значении продуктов этих генов в метаболизме сферохет. В геномах *Sphaerochaetaceae* нет генов секретируемых белков, действующих на первых этапах деградации пектина, а именно пектин- и пектат-лиаз из семейств PL1 и PL9, рамногалактуронан эндо- и экзолиаз семейств PL11 и PL26, полигалактуроназы из семейства GH28 (табл. S1 в приложении).



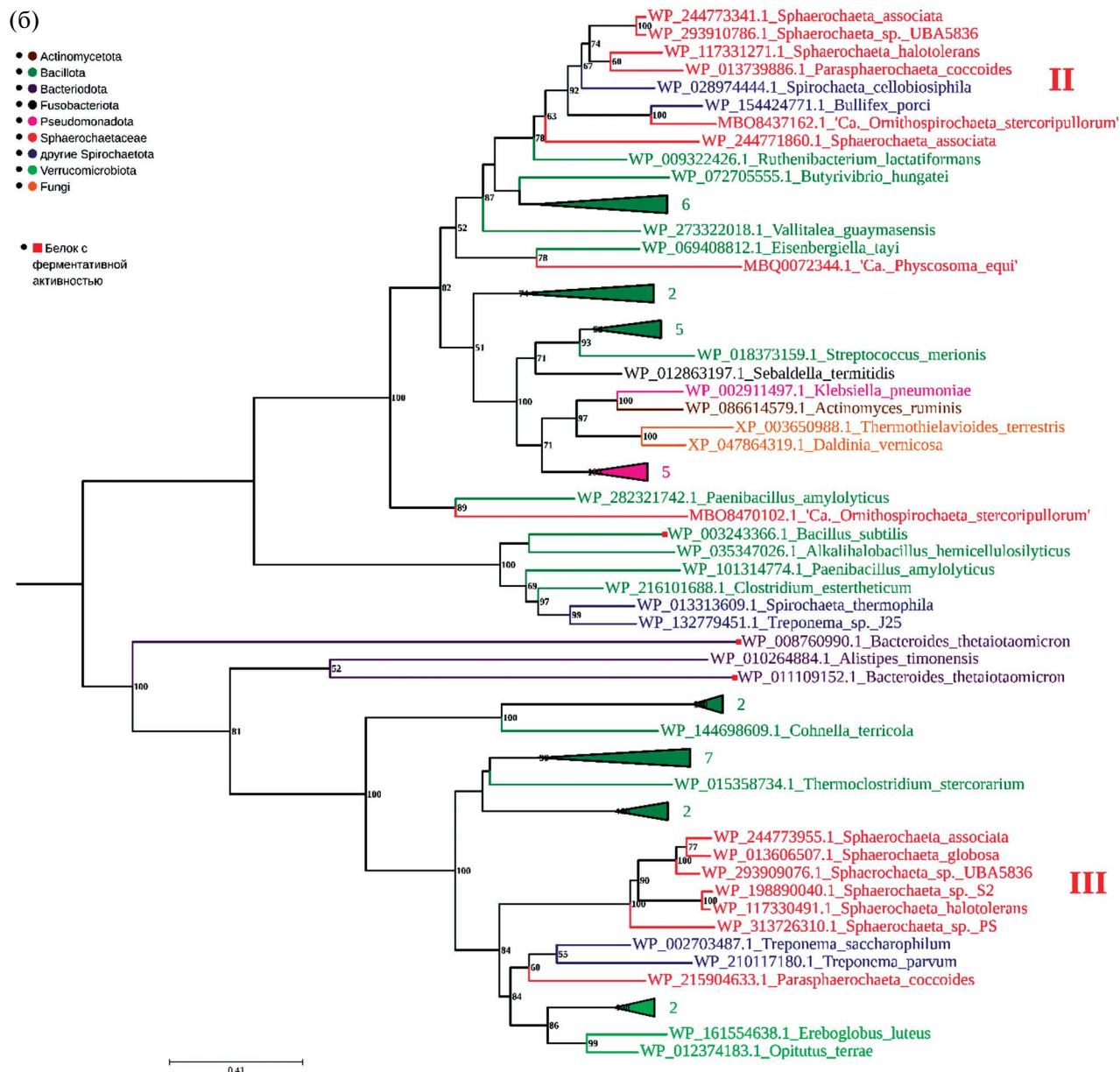


Рис. 1 (а, б). Филогенетическое дерево белков семейства GH105 гликозил-гидролаз, закодированных в геномах *Sphaerochaetaceae*, построенное методом наибольшего правдоподобия: а – рисунок отражает филогенетическое положение белка WP_244772448.1 из *S. associata* и его гомологов; анализ включал 94 аминокислотные последовательности; б – рисунок отражает положение белков WP_244773341.1, WP_244771860.1, WP_244773955.1 из *S. associata* и их гомологов. Анализ включал 74 аминокислотные последовательности. Цветовые обозначения для крупных таксонов указаны на рисунке вверху слева. Белки с определенной ферментативной активностью, а именно d-4,5-ненасыщенная рамногалактуронил-гидролаза (WP_014906487.1, WP_003243366.1, WP_011109152.1) и d-4,5-ненасыщенная α -галактуронидаза (WP_008760990.1, WP_011107571.1, WP_123066362.1) обозначены. В точках ветвления указаны значения бутстрап-поддержки $\geq 50\%$. Масштаб указывает на количество замен на одно положение аминокислотной последовательности. Для упрощения отображения деревьев некоторые кластеры были слопнуты в треугольники со сторонами, соответствующими средней длине ветвей в данном кластере. Количество последовательностей указано рядом с кластерами. Номера последовательностей и организмы в слопнутых кластерах приведены в таблице S3 (по порядку сверху вниз) в приложении.

В то же время в геномах практически всех сферохет есть гены, кодирующие цитоплазматические белки, гомологичные полигалактуроназе (GH28),

пектинметил-эстеразе (CE8), рамногалактуронан-ацетилэстеразе (CE12), ненасыщенной рамногалактуронил-гидролазе (GH105) и α -галактуронидазе

(GH4) (таблица). Последний фермент осуществляет гидролиз дигалактуронатов. Ген потенциальной α -галактуронидазы присутствует у всех исследованных сферохет, за исключением *S. pleomorpha* Grapes^T (таблица).

С целью изучения происхождения белков семейства GH105, кодируемых в геномах *Sphaerochaetaceae*, был проведен филогенетический анализ с использованием нескольких десятков гомологов. В геноме *S. associata* присутствует 4 гена, кодирующих белки семейства GH105 (табл. S1). Поиск ближайших гомологов этих белков в базах NCBI и UniProtKB/Swiss-Prot показал, что один из белков WP_244772448.1 из *S. associata* имел отличающийся спектр близких гомологов и низкий уровень сходства с остальными белками семейства GH105 из *S. associata* (24.9% идентичности с WP_244771860.1, 32.2% с WP_244773341.1 и 30.1% с WP_244773955.1). Кроме того, белок WP_244772448.1 и три остальных белка отличались в отношении их сходства с белками с известной ферментативной активностью. Множественное выравнивание и исследование филогении белков семейства GH105 из *Sphaerochaetaceae* (рис. S3) показало, что WP_244772448.1 и его ближайшие гомологи образуют отдельный и далекий кластер. В связи с низким сходством белка WP_244772448.1 с тремя остальными белками из *S. associata*, для них были построены отдельные филогенетические деревья (рис. 1а и 1б).

На филогенетических деревьях белки семейства GH105 сферохет формируют три четко обособленных кластера I–III (рис. 1а, 1б). Кластер I потенциальных ферментов семейства GH105 сферохет имеет сложную филогенетическую картину (рис. 1а). Попарное сравнение WP_244772448.1 *S. associata* из этого кластера с известными экспериментально охарактеризованными ферментами семейства GH105 показало, что он имеет сходство с d-4,5-ненасыщенной α -галактуронидазой *Paenibacillus amylolyticus* (WP_123066362.1, идентичность 43.2% аминокислотных остатков), с d-4,5-ненасыщенной α -галактуронидазой *B. thetaiotaomicron* (WP_011107571.1, идентичность 43.1%) и ненасыщенной рамногалактуронил-гидролазой YteR из *B. subtilis* (WP_014906487.1, идентичность 36.9%) (Itoh et al., 2006; Ndeh et al., 2017; Keggi, Doran-Peterson 2020). На филогенетическом древе (рис. 1а) кластер I последовательностей сферохет состоит из двух подкластеров. Один из них образован из белков, закодированных в геномах культивируемых изолятов *Sphaerochaetaceae*. Второй образован последовательностями из метабеномов, изолированных из кишечного тракта птиц и млекопитающих. Кроме того, две последовательности из вида “*Ca. Physcosoma equi*” (MBQ0071733.1, MBQ0072301.1), для которого доступен лишь геном, созданный на основе метабенома (MAG),

группируются отдельно от остальных белков сферохет, ближе к последовательностям из *Spirochaetaceae* (рис. 1а). Структура филогенетического древа не позволяет сделать однозначный вывод о происхождении этой группы белков у *Sphaerochaetaceae*, так как кластер I образует статистически плохо поддерживаемые кластеры с белками из *Bacillota*, *Bacterioidota* и некоторыми из *Spirochaetaceae*.

Две последовательности WP_244773341.1 и WP_244771860.1 *S. associata*, вместе с ближайшими гомологами из *Sphaerochaetaceae* и двумя белками из сферохет, образовали кластер II (рис. 1б). Последовательность MBQ0072344.1 из некультивируемого вида “*Ca. Physcosoma equi*” оказалась отдельно от сферохет, вместе с более эволюционно древними белками из *Bacillota*. Белки кластера II имеют ближайшего общего эволюционного предка с представителями *Bacillota* (рис. 1б) и имеют сходство около 30% идентичных аминокислотных остатков с ненасыщенной рамногалактуронил-гидролазой YesR *B. subtilis* (WP_003243366.1) (Itoh et al., 2006).

Кластер III составляют белки длиной около 690 аминокислотных остатков (рис. 1б). Все экспериментально исследованные к настоящему времени белки из семейства GH105 имеют длину до 400 аминокислотных остатков. Попарное сравнение белка WP_244773955.1 *S. associata* с известными охарактеризованными белками семейства GH105 показало низкий уровень сходства с ненасыщенной рамногалактуронил-гидролазой (WP_011109152.1 идентичность 31.8%) и с d-4,5-ненасыщенной α -галактуронидазой (WP_008760990.1, идентичность 27.6%) *B. thetaiotaomicron*. Белки из *Sphaerochaetaceae*, составляющие кластер III, находятся внутри большого статистически надежного кластера последовательностей *Bacillota* (рис. 1б), и, по-видимому, кодирующие их гены были получены в результате горизонтального переноса из *Bacillota*.

Потенциальная α -галактуронидаза семейства GH4 у исследованных сферохет имеет примерно 40% идентичности с гомологами из *Thermoanaerobacterium saccharolyticum* (WP_014757159.1) и *B. subtilis* (WP_003244472.1), для которых была впервые продемонстрирована ферментативная активность — отщепление остатка α -галактуроновой кислоты в дигалактуронатах/олигогалактуронатах (Thompson et al., 2013). Филогенетический анализ потенциальной α -галактуронидазы, кодируемой в геномах сферохет, указывает на общего ближайшего предка у этих белков с белками из веррукомикробиот и протеобактерий (рис. 2).

Исследование роста *S. associata*, *S. globosa*, *S. pleomorpha*, *P. coccoides* показало слабый, но устойчивый рост на пектине в течение многих пересевов. Численность клеток в культурах *S. associata* и *S. globosa* на среде с пектином была значительно выше, чем в контролях (рис. S4 в приложении).

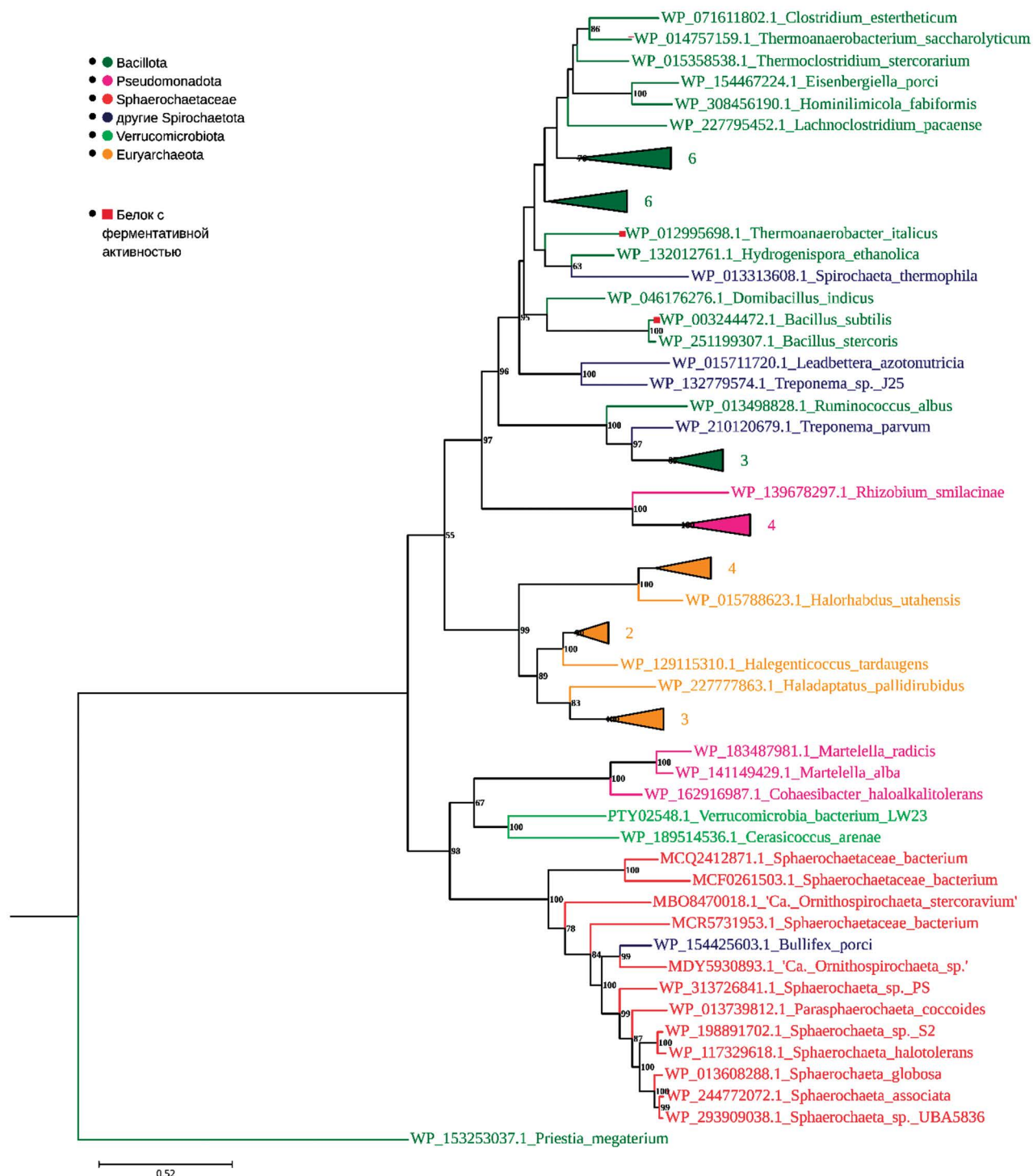


Рис. 2. Филогенетическое древо потенциальных α -галактуронидаз семейства GH4, закодированных в геномах *Sphaerochaetaceae*, построенное методом наибольшего правдоподобия. Анализ включал 67 белковых последовательностей. Значения бутстреп $\geq 50\%$ указаны в точках ветвления. Масштабная линейка указывает на количество замен на одно положение аминокислотной последовательности. Цветовые обозначения для таксонов указаны на рисунке вверху слева. Белки с активностью α -галактуронидазы WP_003244472.1 и WP_012995698.1 обозначены. Для упрощения отображения деревьев некоторые кластеры были схлопнуты в треугольники со сторонами, соответствующими средней длине ветвей в данном кластере. Количество последовательностей указано рядом с кластерами. Номера последовательностей и организмы в схлопнутых кластерах приведены в таблице S3 в приложении.

После 5-го пересева оптическая плотность OD₆₀₀ культуры *S. associata* GLS2^T достигала 0.1 на 11 сут культивирования и дальше не возрастала. Слабый рост бактерий был обусловлен, по-видимому, использованием олигогалактуронидов, присутствующих в коммерческом препарате пектина.

Анализ транспортных систем олигогалактуронидов и путей метаболизма уроновых кислот у сферохет с использованием KEGG (<https://www.kegg.jp/kegg/>) и сервиса BlastKOALA (<https://www.kegg.jp/blastkoala/>) показал, что исследуемые геномы сферохет имеют необходимые наборы генов для транспорта и метаболизма уроновых кислот (табл. S4, S5). Причем *S. associata*, *S. globosa*, *S. halotolerans* 4-11^T, *Sphaerochaeta* sp. S2 имели все потенциальные ферменты для деградации уроновых кислот, в том числе ненасыщенных остатков уроновых кислот, образующихся в результате гидролиза соответствующими гидролазами (семейства GH88 и GH105). В геномах *S. pleomorpha*, *S. halotolerans* 585, *Sphaerochaeta* sp. PS и *P. coccoides* не было лишь одного белка 2-дезоксид-D-глюконат-дегидрогеназы (KduD), работающего на пути превращения ненасыщенных остатков уроновых кислот. Следует отметить, что для метаболизма уроновых кислот в геноме *S. associata* есть дополнительные копии генов, кодирующих алтронат гидролазы (К.Ф. 4.2.1.7), глюкуроноид изомеразы (К.Ф. 5.3.1.12), манноид дегидратазы (К.Ф. 4.2.1.8) (табл. S4). Эти белки имеют в качестве ближайших гомологов последовательности из таксонов *Pseudomonadota* и *Bacillota* и были получены, по-видимому, в результате горизонтальных переносов соответствующих генов. Анализ исследуемых геномов *Sphaerochaetaceae* также выявил многочисленные потенциальные транспортные системы ABC типа для разнообразных компонентов пектина (табл. S5).

В заключение следует отметить, что в геномах исследованных *Sphaerochaetaceae* закодирован умеренный по разнообразию набор ферментов деградации углеводов, среди которых наиболее представлены гены потенциальных α- и β-галактозидаз, α- и β-глюкозидаз, α- и β-ксилозидаз, ненасыщенной рамногалактуронил-гидролазы. Представители исследованных *Sphaerochaetaceae* имеют генетический потенциал для транспорта и гидролиза олигосахаридов разной природы, как растительных, например, компонентов пектина, ксиланов, ксиллоглюканов, так и олигосахаридов микробного происхождения метанохондритина, альгината. Спектр закодированных белков деградации углеводов в незначительной степени коррелировал со спецификой экологической ниши, из которой были выделены эти бактерии. В частности, геном *S. associata* содержал повышенное число генов, продукты которых, вероятно, могут участвовать в деградации метанохондритина, полисахаридного каркаса метаносарцины, из тесной ассоциации

с которой был изолирован штамм *S. associata*. Эти микроорганизмы могут представлять интерес для дальнейшего исследования как источник новых ферментов деградации углеводов. Филогенетический анализ происхождения потенциальных α-галактуронидаз и ненасыщенной рамногалактуронил-гидролазы, закодированных в геномах *Sphaerochaetaceae*, указывает на их приобретение от общего предка с *Pseudomonadota* (GH4) и предков разного таксономического положения для белков семейства GH105.

СОБЛЮДЕНИЕ ЭТИЧЕСКИХ СТАНДАРТОВ

Настоящая статья не содержит результатов исследований с использованием животных в качестве объектов.

КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют об отсутствии конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

- Березина О.В., Лунина Н.А., Зверлов В.В., Наумов Д.Г., Либель В., Великодворская Г.А. Кластер генов *Thermotoga neopolitana*, участвующих в деградации крахмала и мальтодекстринов: молекулярная структура локуса // Мол. биология. 2003. Т. 37. С. 801–809.
- Berezina O. V., Lunina N. A., Zverlov V. V., Naumoff D. G., Liebl W., Velikodvorskaya G. A. A cluster of *Thermotoga neopolitana* genes involved in the degradation of starch and maltodextrins: the molecular structure of the locus // Mol. Biol. (Moscow). 2003. V. 37. P. 678–685.
- Кевбрин В.В., Заварзин Г.А. Влияние соединений серы на рост галофильной гомоацетатной бактерии *Acetohalobium arabaticum* // Микробиология. 1992. Т. 61. С. 812–817.
- Kevbrin V. V., Zavarzin G. A. The effect of sulfur compounds on growth of halophilic homoacetic bacterium *Acetohalobium arabaticum* // Microbiology (Moscow). 1992. V. 61. P. 563–571.
- Наумов Д.Г. Иерархическая классификация гликозил-гидролаз // Биохимия. 2011. Т. 76. С. 764–780.
- Naumoff D. G. Hierarchical classification of glycoside hydrolases // Biochemistry (Moscow). 2011. V. 76. P. 622–635.
- Abt B., Han C., Scheuner C., Lu M., Lapidus A., Nolan M., Lucas S., Hammon N., Deshpande S., Cheng J.-F., Tapia R., Goodwin L. A., Pitluck S., Liolios K., Paganini I., Ivanova N., Mavromatis K., Mikhailova N., Huntemann M., Pati A., Chen A., Palaniappan K., Land M., Hauser L., Brambilla E.-M., Rohde M.,

- Spring S., Gronow S., Goeker M., Woyke T., Bristow J., Eisen J.A., Markowitz V., Hugenholtz P., Kyrpides N.C., Klenk H.-P., Detter J.C. Complete genome sequence of the termite hindgut bacterium *Spirochaeta coccoides* type strain (SPN1^T), reclassification in the genus *Sphaerochaeta* as *Sphaerochaeta coccoides* comb. nov. and emendations of the family *Spirochaetaceae* and the genus *Sphaerochaeta* // Stand. Genom. Sci. 2012. V. 6. P. 194–209.
- Arnold B.J., Huang I.T., Hanage W.P. Horizontal gene transfer and adaptive evolution in bacteria // Nat. Rev. Microbiol. 2022. V. 20. P. 206–218.
- Bianchini G., Sánchez-Baracaldo P. TreeViewer: Flexible, modular software to visualise and manipulate phylogenetic trees // Ecol. Evol. 2024. V. 14. Art. e10873.
- Caro-Quintero A., Deng J., Auchtung J., Brettar I., Höfle M.G., Klappenbach J., Konstantinidis K.T. Unprecedented levels of horizontal gene transfer among spatially co-occurring *Shewanella* bacteria from the Baltic Sea // ISME J. 2011. V. 5. P. 131–140.
- Caro-Quintero A., Ritalahti K.M., Cusick K.D., Löffler F.E., Konstantinidis K.T. The chimeric genome of *Sphaerochaeta*: nonspiral spirochetes that break with the prevalent dogma in spirochete biology // MBio. 2012. V. 3. Art. e00025-12.
- Chuvochina M., Mussig A.J., Chaumeil P.A., Skarszewski A., Rinke C., Parks D.H., Hugenholtz P. Proposal of names for 329 higher rank taxa defined in the Genome Taxonomy Database under two prokaryotic codes // FEMS Microbiol. Lett. 2023. V. 370. Art. fnad071.
- Drula E., Garron M.L., Dogan S., Lombard V., Henrissat B., Terrapon N. The carbohydrate-active enzyme database: functions and literature // Nucl. Acids Res. 2022. V. 50. P. D571–D577.
- Frank K., Sippl M.J. High-performance signal peptide prediction based on sequence alignment techniques // Bioinformatics. 2008. V. 24. P. 2172–2176.
- Gharechahi J., Sarikhan S., Han J.L., Ding X.Z., Sal-ekdeh G.H. Functional and phylogenetic analyses of camel rumen microbiota associated with different lignocellulosic substrates // NPJ Biofilms Microbiomes. 2022. V. 8. Art. 46.
- Gilchrist C.L., Booth T.J., van Wersch B., van Grieken L., Medema M.H., Chooi Y.H. cblaster: a remote search tool for rapid identification and visualization of homologous gene clusters // Bioinform. Adv. 2021. V. 1. Art. vbab016.
- Gilroy R., Leng J., Ravi A., Adriaenssens E.M., Oren A., Baker D., La Ragione R.M., Proudman C., Pallen M.J. Metagenomic investigation of the equine faecal microbiome reveals extensive taxonomic diversity // PeerJ. 2022. V. 10. Art. 13084.
- Gilroy R., Ravi A., Getino M., Pursley I., Horton D.L., Alikhan N.F., Baker D., Gharbi K., Hall N., Watson M., Adriaenssens E.M., Foster-Nyarko E., Jarju S., Sec-ka A., Antonio M., Oren A., Chaudhuri R.R., La Ra-gione R., Hildebrand F., Pallen M.J. Extensive microbial diversity within the chicken gut microbiome revealed by metagenomics and culture // PeerJ. 2021. V. 9. Art. 10941.
- Grouzdev D.S., Bidzhieva S.K., Sokolova D.S., Tourova T.P., Patutina E.O., Poltarau A.B., Nazina T.N. Draft genome sequence of a fermenting bacterium, “*Sphaerochaeta halotolerans*” 4-11^T, from a low-temperature petroleum reservoir in Russia // Microbiol. Res. Announc. 2018. V. 7. Art. e01345-18.
- Hall B.G., Pikis A., Thompson J. Evolution and biochemistry of family 4 glycosidases: implications for assigning enzyme function in sequence annotations // Mol. Biol. Evol. 2009. V. 26. P. 2487–2497.
- Itoh T., Ochiai A., Mikami B., Hashimoto W., Murata K. A novel glycoside hydrolase family 105: the structure of family 105 unsaturated rhamnogalacturonyl hydrolase complexed with a disaccharide in comparison with family 88 enzyme complexed with the disaccharide // J. Mol. Biol. 2006. V. 360. P. 573–585.
- Kanehisa M., Sato Y., Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences // J. Mol. Biol. 2016. V. 428. P. 726–731.
- Keggi C., Doran-Peterson J. The homogalacturonan deconstruction system of *Paenibacillus amylolyticus* 27C64 requires no extracellular pectin methylesterase and has significant industrial potential // Appl. Environ. Microbiol. 2020. V. 86. Art. e02275-19.
- Keggi C., Doran-Peterson J. *Paenibacillus amylolyticus* 27C64 has a diverse set of carbohydrate-active enzymes and complete pectin deconstruction system // J. Ind. Microbiol. Biotechnol. 2019. V. 46. P. 1–11.
- Klingl A., Pickl C., Flechsler J. Archaeal cell walls // Sub-cell Biochem. / Ed. Kuhn A. Springer Cham., 2019. V. 92. P. 471–493.
https://doi.org/10.1007/978-3-030-18768-2_14
- Morrison P.K., Newbold C.J., Jones E., Worgan H.J., Grove-White D.H., Dugdale A.H., Barfoot C., Harris P.A., Argo C.M. The equine gastrointestinal microbiome: impacts of weight-loss // BMC Vet. Res. 2020. V. 16. P. 78.
- Müller M., Calvert M., Hottmann I., Kluj R.M., Teufel T., Balbucht K., Engelbrecht A., Selim K.A., Xu Q., Borisova M., Titz A., Mayer C. The exo-β-N-acetylmuramidase NamZ from *Bacillus subtilis* is the founding member of a family of exo-lytic peptidoglycan hexosaminidases // J. Biol. Chem. 2021. V. 296. Art. 100519.
- Nazina T.N., Bidzhieva S.K., Grouzdev D.S., Löffler F.E. *Sphaerochaeta* // Bergey’s Manual of Systematics of Archaea and Bacteria / Eds. Trujillo M.E., Dedysh S., DeVos P., Hedlund B., Kämpfer P., Rainey F.A., Whitman W.B. John Wiley & Sons, Inc., in association with Bergey’s Manual Trust, 2021. P. 1–13.
<https://doi.org/10.1002/9781118960608.gbm02038>
- Nazina T.N., Bidzhieva S.K., Grouzdev D.S., Löffler F.E. *Sphaerochaetaceae* // Bergey’s Manual of Systematics of Archaea and Bacteria / Eds. Trujillo M.E., Dedysh

- S., DeVos P., Hedlund B., Kämpfer P., Rainey F.A., Whitman W.B. John Wiley & Sons, Inc., in association with Bergey's Manual Trust, 2022. P. 1–6.
<https://doi.org/10.1002/9781118960608.fbm00390>
- Ndeh D., Rogowski A., Cartmell A., Luis A.S., Baslé A., Gray J., Venditto I., Briggs J., Zhang X., Labourel A., Terrapon N., Buffetto F., Nepogodiev S., Xiao Y., Field R.A., Zhu Y., O'Neil M.A., Urbanowicz B.R., York W.S., Davies G.J., Abbott D.W., Ralet M.C., Martens E.C., Henrissat B., Gilbert H.J. Complex pectin metabolism by gut bacteria reveals novel catalytic functions // *Nature*. 2017. V. 544. P. 65–70.
- Nguyen P., Eshaque R., Garland B.A., Dang A., Suits M.D.L. Degradation of chondroitin sulfate A by a PUL-like operon in *Tannerella forsythia* // *PLoS One*. 2022. V. 17. Art. e0272904.
- O'Sullivan O., O'Callaghan J., Sangrador-Vegas A., McAuliffe O., Slattery L., Kaleta P., Callanan M., Fitzgerald G.F., Ross R.P., Beresford T. Comparative genomics of lactic acid bacteria reveals a niche-specific gene set // *BMC Microbiol.* 2009. V. 9. Art. 50.
- Ochman H., Lawrence J.G., Groisman E.A. Lateral gene transfer and the nature of bacterial innovation // *Nature*. 2000. V. 405. P. 299–304.
- Tamura K., Stecher G., Kumar S. // MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol. Biol. Evol.* 2021. V. 38. P. 3022–3027.
- Tang S., Xin Y., Ma Y., Xu X., Zhao S., Cao J. Screening of microbes associated with swine growth and fat deposition traits across the intestinal tract // *Front. Microbiol.* 2020. V. 11. Art. 586776.
- Thompson J., Pikis A., Rich J., Hall B.G., Withers S.G. α -Galacturonidase(s): a new class of family 4 glycoside hydrolases with strict specificity and a unique CHEV active site motif // *FEBS Lett.* 2013. V. 587. P. 799–803.
- Troshina O., Oshurkova V., Suzina N., Machulin A., Ariskina E., Vinokurova N., Kopitsyn D., Novikov A., Shcherbakova V. *Sphaerochaeta associata* sp. nov., a spherical spirochaete isolated from cultures of *Methanosarcina mazei* JL01 // *Int. J. Syst. Evol. Microbiol.* 2015. V. 65. P. 4315–4322.
- Troshina O.Y., Tepeeveva A.N., Arzamastseva V.O., Whitman W.B., Varghese N.J., Shapiro N., Woyke T., Kyrpides N.C., Vasilenko O.V. Complete genome sequence of the type strain bacterium *Sphaerochaeta associata* GLS2^T (VKM B-2742)^T // *J. Bioinform. Genom.* 2024. V. 1. Art. 23.
<http://dx.doi.org/10.18454/jbg.2024.23.1>
- Ulmer J.E., Vilén E.M., Namburi R.B., Benjdia A., Beneteau J., Malleron A., Bonnaffé D., Dri-guez P.A., Descroix K., Lassalle G., Le Narvor C., Sandström C., Spillmann D., Berteau O. Characterization of glycosaminoglycan (GAG) sulfatases from the human gut symbiont *Bacteroides thetaiotaomicron* reveals the first GAG-specific bacterial endosulfatase // *J. Biol. Chem.* 2014. V. 289. P. 24289–24303.
- Zheng J., Ge Q., Yan Y., Zhang X., Huang L., Yin Y. db-CAN3: automated carbohydrate-active enzyme and substrate annotation // *Nucl. Acids Res.* 2023. V. 51. P. W115–W121.

EXPERIMENTAL ARTICLES

Comparative Genomics of Carbohydrates Utilization in Bacteria of the Family *Sphaerochaetaceae*: Evolutionary Origin of the Genes Encoding Galacturonidase and Unsaturated Rhamnogalacturonyl Hydrolase

O. Yu. Troshina¹*, D. G. Naumoff², V. I. Rechkina¹, and V. A. Shcherbakova¹

¹Scryabin Institute of Biochemistry and Physiology of Microorganisms,
Pushchino Scientific Center for Biological Research, Russian Academy of Sciences, Pushchino, 142290, Russia

²Winogradsky Institute of Microbiology, Research Centre of Biotechnology, Russian Academy of Sciences,
Moscow, 119071, Russia

*e-mail: oltro676@yandex.ru

Abstract. A comparative analysis of carbohydrate degradation proteins encoded in currently available genomic sequences of bacteria of the family *Sphaerochaetaceae*, namely *Sphaerochaeta associata* GLS2^T, *S. globosa* Buddy^T, *S. pleomorpha* Grapes^T, *S. halotolerans* 4-11^T, *S. halotolerans* 585, *Sphaerochaeta* sp. S2, *Sphaerochaeta* sp. PS and *Parasphaerochaeta coccoides* SPN1^T was carried out. The genomes of *Sphaerochaeta* spp. encode a medium-sized and diverse set of proteins potentially involved in the degradation of different classes of carbohydrates, mainly oligosaccharides. All studied genomes encode glycoside hydrolases of the GH1, GH2, GH3, GH4, GH13, GH20, GH28, GH36, GH43, GH57, GH63, GH77 and GH105 families, as well as carbohydrate esterases of the CE8 and

CE9 families. All studied bacteria, with the exception of *P. coccoides* SPN1^T, have many proteins of the GH31 family encoded in their genomes. The studied representatives of *Sphaerochaetaceae* do not have genes coding for endo- β -acetylmuramidase (lysozyme) of the GH23 family involved in the process of peptidoglycan turnover. However, the genomes of *S. associata*, *S. globosa*, *Sphaerochaeta* sp. PS and *S. pleomorpha* contain the exo- β -acetylmuramidase gene (GH171 family). A significant part of the genes encoding carbohydrate degradation enzymes have the closest homologues among representatives of the phyla *Bacillota*, *Bacteroidota*, and *Pseudomonadota*. The genomes of the studied bacteria encode proteins that could potentially be involved in the degradation of pectin. The ability of representatives of *Sphaerochaetaceae* to use pectin for growth, as well as the evolutionary origin of genes encoding potential α -galacturonidase (GH4 family) and unsaturated glucuronyl/rhamnogalacturonyl hydrolase (GH105 family), involved in the degradation of pectin components, were studied.

Keywords: *Sphaerochaeta*, lateral gene transfer, glycoside hydrolase, polysaccharide lyase, CAZy, GH4, GH105, α -galacturonidase, unsaturated glucuronyl/rhamnogalacturonyl hydrolase