

На правах рукописи

МАРЧЕНКОВ Артём Михайлович

**СИСТЕМА ГЕНОВ ТРАНСПОРТА КРЕМНИЕВОЙ КИСЛОТЫ
У ПРЕДСТАВИТЕЛЕЙ ДИАТОМОВЫХ ВОДОРОСЛЕЙ РОДА**

SYNEDRA EHRENBURG

03.01.07 – молекулярная генетика

Автореферат
диссертации на соискание учёной степени
кандидата биологических наук

Новосибирск – 2019

Работа выполнена в отделе ультраструктуры клетки Федерального государственного бюджетного учреждения науки Лимнологический институт Сибирского отделения Российской академии наук, г. Иркутск.

Научный
руководитель:

Грачев Михаил Александрович, доктор химических наук, академик РАН, главный научный сотрудник, Федеральное государственное бюджетное учреждение науки Лимнологический институт Сибирского отделения Российской академии наук, г. Иркутск.

Официальные
оппоненты:

Марданов Андрей Владимирович, доктор биологических наук, профессор РАН, заведующий лабораторией, Федеральное государственное учреждение «Федеральный исследовательский центр «Фундаментальные основы биотехнологии» Российской академии наук», г. Москва.

Ершов Никита Игоревич, кандидат биологических наук, старший научный сотрудник, Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», г. Новосибирск.

Ведущая организация:

Федеральное государственное бюджетное учреждение науки Институт биологии гена Российской академии наук, г. Москва.

Защита состоится: _____ 2019 г. в ____ часов на заседании диссертационного совета Д 003.074.01 по защите диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук при ФГБУН Институт молекулярной и клеточной биологии СО РАН (630090, Новосибирск, пр. академика Лаврентьева, 8/2, тел. (383)-373-02-49, e-mail: ovant@mcb.nsc.ru). С диссертацией можно ознакомиться в библиотеке и на сайте Института молекулярной и клеточной биологии СО РАН https://www.mcb.nsc.ru/diss_council/autoref

Автореферат разослан «____» _____ 2019 г.

Ученый секретарь
диссертационного совета,
кандидат биологических наук

О.В. Антоненко

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность проблемы

Диатомовые водоросли (Bacillariophyta) – это одна из самых многочисленных и экологически важных групп фитопланктона. Диатомеи обеспечивают 20 % первичной продукции органического углерода Земли [Tréguer *et al.*, 1995; Tréguer *et al.*, 2018] и играют ведущую роль в биогеохимическом цикле кремния [Nelson *et al.*, 1995]. Кремний используется диатомеями для строительства видоспецифически орнаментированных экзоскелетов из кремнезема. Биосилификация встречается и у других организмов, относящихся к разным таксонам [Müller *et al.*, 2003]. Одной из ключевых проблем биосилификации являются механизмы захвата кремния из окружающей среды и его транспорта внутрь клетки. Наибольший прогресс в изучении данного процесса достигнут у диатомей, хотя многие вопросы остаются неизвестными. В 1997 г. было открыто семейство из пяти генов, кодирующих белки транспорта кремния (silicon transporter, SIT) у морской пеннатной шовной диатомеи *Cylindrotheca fusiformis* Reimann *et al.* [Hildebrand *et al.*, 1997; Hildebrand, Dahlin, Volcani, 1998]. Второй диатомеей, у которой был обнаружен ген *sit*, была *Synedra acus* [Грачев и др., 2002]. К настоящему времени расшифровано свыше 400 генов *sit* из более чем 100 видов диатомей [Durkin *et al.*, 2016]. Зачастую в геномах диатомей имеются по несколько паралогичных генов протяженностью 1500-1800 пар нуклеотидов (п.н.) [Hildebrand, Dahlin, Volcani, 1998; Thamtrakoln, Alverson, Hildebrand, 2006; Sapriel *et al.*, 2009]. Гены *sit* идентифицированы у широкого круга организмов из разных таксонов: хризофитовых [Лихошвай и др., 2006], хоанофлагеллят [Marron *et al.*, 2013], гаптофитовых водорослей [Durak *et al.*, 2016], веслоногого рачка *Calanus finmarchicus* Gunnerus, многощетинковых червей, ризарий, радиолярий, инфузорий и цианобактерий [Durkin *et al.*, 2016; Marron *et al.*, 2016]. Белки SIT имеют 10 трансмембранных доменов (ТМ) [Hildebrand, 2000; Thamtrakoln, Hildebrand, 2005; Sapriel *et al.*, 2009; Durkin *et al.*, 2016], один консервативный мотив CMLD [Грачев и др., 2002; Щербакова и др., 2005; Grachev *et al.*, 2005] и четыре консервативных мотива GXQ [Thamtrakoln, Alverson, Hildebrand, 2006].

В результате анализа данных транскриптома диатомеи *Pseudo-nitzschia multiseries* (Hasle) Hasle впервые был обнаружен ген *sit* (мульти-*sit*) длиной 4548 п.н. Соответствующая ему аминокислотная последовательность имеет три слитых друг с другом «зрелых» белка

SIT [Durkin *et al.*, 2012]. Под предсказанным «зрелым» белком SIT подразумевается внутренняя наиболее консервативная часть белка SIT, содержащая все высококонсервативные аминокислотные остатки (а.о.), которые присутствуют у белков SIT и могут принимать непосредственное участие в транспорте кремниевой кислоты через мембрану.

В результате аннотации данных массового параллельного секвенирования ядерного генома *Synedra acus* subsp. *radians* (Kütz.) Skabitsch., полученных на платформах 454/Roche GS FLX и Illumina MiSeq, был идентифицирован ген *sit2* [Петрова, 2012]. Так как открытая рамка считывания (ОРС) гена *sit2* была незавершенной на 5'-конце, стартовый кодон отсутствовал, это свидетельствовало о том, что данная последовательность не является полноразмерным геном. Для дальнейших структурно-функциональных исследований белков SIT необходимы данные о полноразмерных последовательностях генов, кодирующих данные белки.

Цель настоящей работы – определение полноразмерных последовательностей мульти-*sit* генов у пресноводных пеннатных диатомовых водорослей *S. acus* subsp. *radians* и *Synedra ulna* subsp. *danica*.

Были поставлены следующие **задачи**:

1. Секвенировать полноразмерные нуклеотидные последовательности мульти-*sit* генов у диатомей *S. acus* subsp. *radians* и *S. ulna* subsp. *danica*.
2. Установить взаимное расположение мульти-*sit* генов в геномах *S. acus* subsp. *radians* и *S. ulna* subsp. *danica*.
3. Провести сравнительный анализ белков SIT, состоящих из нескольких слитых друг с другом предсказанных «зрелых» белков.

Научная новизна. Впервые определены нуклеотидные последовательности генов мульти-*sit* у пресноводных пеннатных бесшовных диатомовых водорослей *Synedra acus* subsp. *radians* и *Synedra ulna* subsp. *danica*. Впервые показано, что гены мульти-*sit* могут располагаться в одной хромосоме.

Практическая и теоретическая значимость работы. Полученные данные расширяют представления о генах *sit* у диатомовых водорослей. Полученные последовательности мульти-*sit* генов зарегистрированы в базе данных GenBank и могут быть использованы для поиска генов *sit* у других силифицирующих организмов и определения путей их эволюции. Полученные результаты

о нуклеотидных последовательностях мульти-*sit* генов позволяют планировать проведение дальнейших исследований по их экспрессии в гетерологичной системе и экспериментов по идентификации и определению первичной структуры соответствующих белков в протеомах.

Основные положения диссертации, выносимые на защиту:

1. Мульти-*sit* гены пресноводных пениатных диатомей рода *Synedra* относятся к новой малоизученной группе генов *sit* и потенциально способны кодировать слитые белки SIT, состоящие из нескольких «зрелых» белков.
2. В геноме диатомеи *Synedra acus* subsp. *radians* два мульти-*sit* гена располагаются в одной хромосоме. В геноме *Synedra ulna* subsp. *danica* – три мульти-*sit* гена, два из которых полностью идентичны и располагаются в одной хромосоме.
3. Мульти-SIT белки состоят из 2 или 3 слитых друг с другом «зрелых» белков SIT, на границах которых расположены консервативные мотивы (DXDID), которые являются гипотетическими сайтами протеолиза при посттрансляционном созревании белка.

Личный вклад автора. Автор принимал непосредственное участие в планировании и проведении работы. Все результаты за исключением получения моноклональных культур диатомовых водорослей получены лично автором либо при его непосредственном участии в ходе коллективных работ. Статьи по результатам проведенных работ подготовлены в соавторстве и опубликованы в рецензируемых изданиях.

Апробация работы. Материалы диссертации были представлены на VI-ом Всероссийском с международным участием конгрессе молодых ученых «Симбиоз России 2013» (Иркутск, 2013), VII-ой Международной школе молодых учёных по молекулярной генетике «Геномика и биология живых систем» (Москва, 2016), IV-ой Международной конференции молодых ученых: биотехнологов, молекулярных биологов и вирусологов (Новосибирск, 2017).

Публикации. По материалам диссертационной работы опубликовано 3 научные статьи в изданиях из Перечня ВАК РФ, одна статья в научно-популярном журнале и 3 тезиса в материалах конференций.

Структура и объём работы. Диссертация состоит из введения, обзора литературы, описания материалов и методов, результатов и

обсуждения, а также выводов и списка цитируемой литературы, в которые входит 125 ссылок. Работа изложена на 154 страницах машинописного текста и содержит 29 иллюстраций, 4 таблицы и 11 приложений.

СОДЕРЖАНИЕ РАБОТЫ МАТЕРИАЛЫ И МЕТОДЫ

Объекты исследований. Объектами исследования были пеннатные бесшовные виды диатомей, обитающие в оз. Байкал. По мнению ряда авторов, *S. acus* subsp. *radians* является синонимом *Fragilaria radians* (Kütz.) D.M. Williams & Round [Williams, Round, 1987], а *S. ulna* subsp. *danica* – синонимом *Ulnaria danica* (Kütz.) Compère & Bukhtiyarova [Bukhtiyarova, Compère, 2006].

Исследованные в работе таксоны имеют высокое морфологическое сходство, были определены на основе диагнозов, приведенных в монографии Г.И. Поповской, С.И. Генкала и Е.В. Лихошвай [2011] и введены в моноклональную культуру группой культивирования диатомовых водорослей к.б.н. Ю.Р. Захаровой и Н.А. Волокитиной (отдел ультраструктуры клетки, ЛИН СО РАН). Клетки *S. acus* subsp. *radians* и *S. ulna* subsp. *danica* культивировали на среде DM (Diatom Medium) [Thompson *et al.*, 1988] в течение четырех недель.

Выделение ДНК. Высокомолекулярную фракцию геномной ДНК из биомассы *S. acus* subsp. *radians* и *S. ulna* subsp. *danica* выделяли по описанному ранее методу [Jacobs *et al.*, 1992] с незначительными модификациями.

Идентификация мульти-*sit* генов *S. acus* subsp. *radians*. Для идентификации мульти-*sit* генов *S. acus* subsp. *radians* использовали как данные массового параллельного секвенирования (Next Generation Sequencing, NGS) генома, так и секвенирование по методу Сэнгера.

Идентификация мульти-*sit* генов *S. ulna* subsp. *danica*. Для идентификации мульти-*sit* использовали вырожденные разнаправленные праймеры (UniCMLDF – UniCMLDR), которые были выбраны на основе сравнения участков генов *sit*, кодирующих мотивы CMLD и их окружение. Полученные ПЦР-продукты клонировали и секвенировали в составе плазмидной ДНК. Связь между расшифрованными ПЦР фрагментами идентифицировали, используя праймеры, расположенные в разных ПЦР-продуктах.

Выделение РНК и RACE-ПЦР. Суммарную РНК из клеток выделяли с помощью набора RNeasy Plant Mini Kit (“QIAGEN”, Германия). 5'- и 3'- концы мРНК генов мульти-*sit* определяли, используя набор Mint RACE cDNA amplification set («Евроген», Россия).

Анализ нуклеотидных и аминокислотных последовательностей. Выравнивание нуклеотидных и предсказанных аминокислотных последовательностей относительно друг друга проводили с помощью алгоритма CLUSTALW в программе BioEdit 7.2.5 [Hall, 1999] и Vector NTI 11. Поиск трансмембранных доменов проводили при помощи онлайн программы TmPred [http://www.ch.embnet.org/software/TMPRED_form.html]. Границы предсказанных «зрелых» белков SIT в последовательностях определяли посредством онлайн-поиска соответствующей модели PF03842 в базе данных Pfam [<http://pfam.xfam.org/>] [Finn *et al.*, 2016].

Филогенетический анализ. Матрицы замен НКУ [Kishino, Hasegawa, 1989] и LG [Le, Gascuel, 2008] выбирали при помощи функции “find best DNA/Protein models (ML)”, пакет программ MEGA 6.06 [Tamura *et al.*, 2013]. Филогенетический анализ методом максимального правдоподобия (Maximum Likelihood Method) проводился пакетом программ MEGA 6.06 [Tamura *et al.*, 2013]. Для оценки поддержки было построено 1000 бутстрэп-деревьев [Zuckerkandl, Pauling, 1965]. Байесовский филогенетический анализ проводили на том же наборе данных с помощью MrBayes 3.2.6 [Ronquist, Huelsenbeck, 2003]. Анализировали 16 МСМCMC-цепей в течение 600000 поколений, модель замен не фиксировалась.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Исследование генов мульти-*sit* у диатомеи *S. acus* subsp. *radians*

В результате секвенирования генома *S. acus* subsp. *radians* нами был получен контиг axSAfrepc37183 протяженностью 5021 п.н., имеющий гомологию с известными последовательностями *sit*. При анализе контига axSAfrepc37183 показано, что его 5'-конец (позиции 423-1864 п.н.) на 97,8 % идентичен гену *sit1* (GenBank FJ263063) *S. acus* subsp. *radians*, а середина (позиции 1288-3000 п.н.) на 100 % идентична гену *sit2* (GenBank JQ595302, **рис. 1А**).

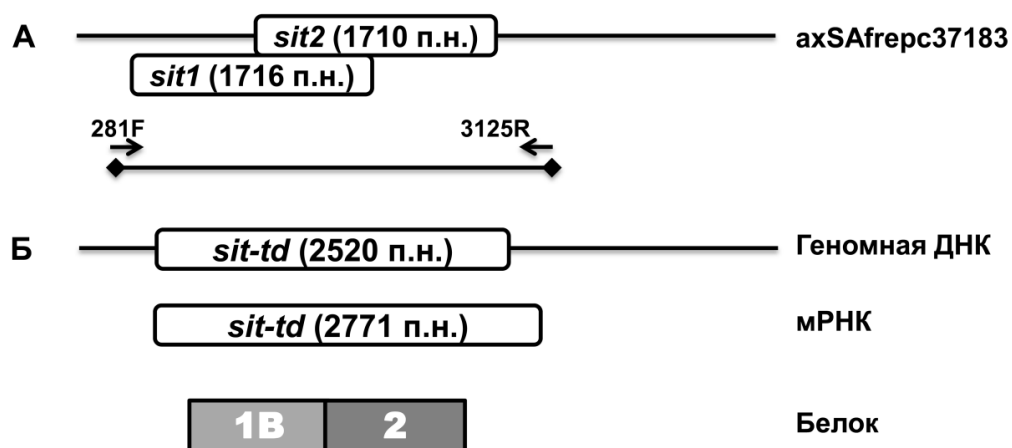


Рисунок 1. Схема эксперимента по получению гена *sit-td* *S. acus* subsp. *radians*. А – контиг axSAfrepc37183, на котором отмечены места выравнивания генов *sit1* и *sit2*. Б – ОПС гена *sit-td*, мРНК гена *sit-td* и предсказанная аминокислотная последовательность SIT-TD. Прямоугольники – ОПС *sit1*, *sit2* и *sit-td*. 1В и 2 – предсказанные «зрелые» белки SIT, границы которых определены посредством онлайн-поиска доменов в базе данных Pfam. → – места посадки праймеров.

Для уточнения последовательности контига axSAfrepc37183 была использована сконструированная пара праймеров 281F-3125R и получен ПЦР-продукт длиной 2844 п.н. (рис. 1А). Сравнение контига axSAfrepc37183 с полученными данными секвенирования по методу Сэнгера показало, что между нуклеотидными последовательностями существует 3 отличия (инсерции T959 и G982; замена 1004T>C; позиции указаны по контигу axSAfrepc37183). В результате исправления ошибок сборки ОПС гена *sit2* была увеличена с 1710 п.н. до 2520 п.н. и ген был назван *sit-td* (рис. 1Б). Предсказанная аминокислотная последовательность белка SIT-TD *S. acus* subsp. *radians* состоит из двух слитых предсказанных «зрелых» белков SIT 1В (392 а.о.) и SIT 2 (441 а.о.), содержит два мотива CMLD и восемь мотивов GXQ (рис. 1Б).

Парное выравнивание нуклеотидных последовательностей фрагмента гена *sit1* и контига, содержащего ген *sit-td*, показало, что 5'-конец (272 п.н.) гена *sit1* на 43,2 % идентичен некодирующему участку ДНК, расположенному перед 5'- концом гена *sit-td* (рис. 2).

При проведении ПЦР с использованием прямого праймера 10F на 5'- область *sit1* и обратного праймера 3125R (рис. 3А), комплементарного участку гена *sit-td*, синтезировался фрагмент протяженностью 2976 п.н., 3'- область которого идентична *sit-td* (рис. 3Б). С использованием данных NGS генома *S. acus* subsp. *radians*

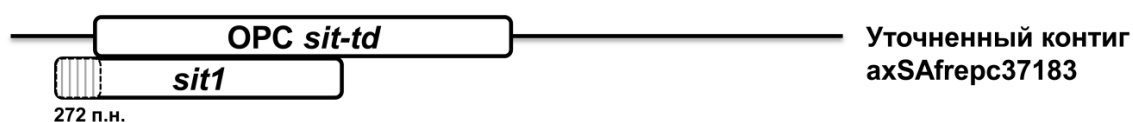


Рисунок 2. Схема перекрытия уточненного контига axSAfrepc37183 и гена *sit1*. Вертикальной штриховкой отмечен 5'-конец (272 п.н.) гена *sit1*, который на 43,2 % идентичен некодирующему участку перед ОПС гена *sit-td*.

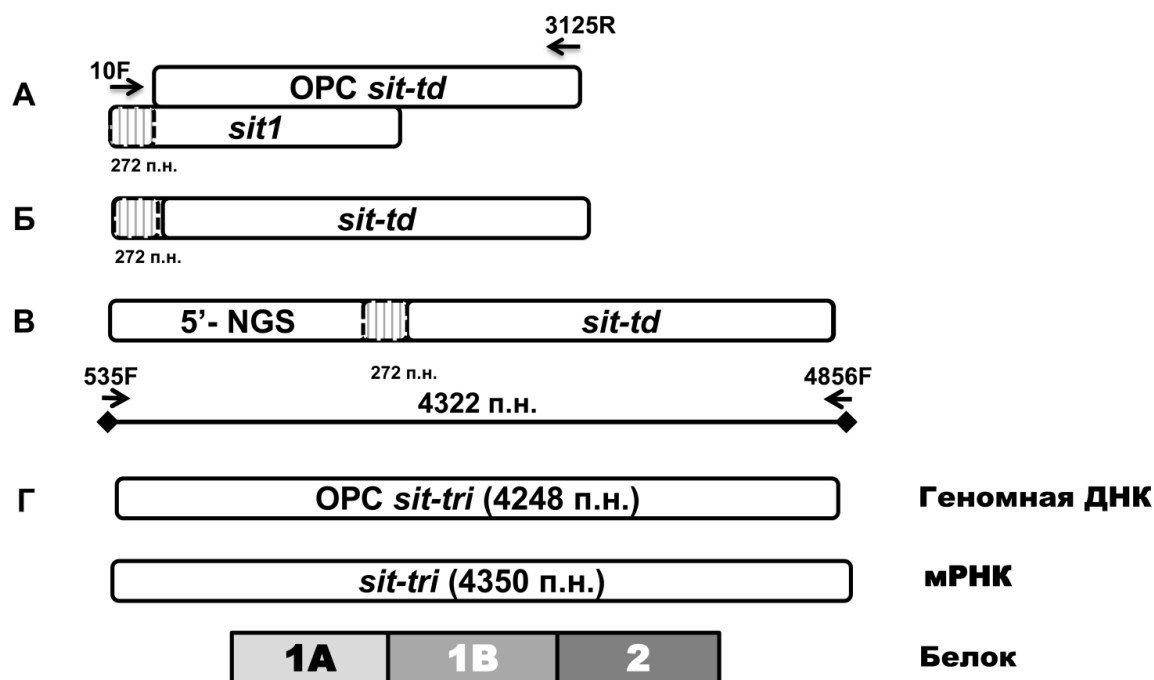


Рисунок 3. Схема эксперимента по определению гена *sit-tri* *S. acus* subsp. *radians*. А – сравнение гена *sit-td* и фрагмента гена *sit1*, вертикальной штриховкой отмечен участок гена *sit1* (272 п.н.), отличающийся от *sit-td*. Б – в результате секвенирования показано, что полученный продукт на 5'-конце имеет участок из 272 п.н., имеющий 100 % идентичность гену *sit1*, далее следует ген *sit-td*. В – «ручная» сборка 5'-конца фрагмента гена *sit1*'. Г – ОПС гена *sit-tri*. → – места посадки праймеров.

Хабудаевым К.В. (ЛИН СО РАН) была проведена «ручная» сборка 5'-области *sit1*' и получен контиг протяженностью 1505 п.н. В результате объединения контигов, полученных с помощью секвенирования по методу Сэнгера и NGS, длина контига была увеличена до 4422 п.н., а ОПС составила 4281 п.н. (**рис. 3В**).

Используя соответствующие контигу пары праймеров 535F-4856R, был получен ПЦР-продукт (4322 п.н.) (**рис. 3В**). Далее полученный ПЦР-продукт был секвенирован с помощью праймеров: 10F, 1210R, 2298R, 2524R, 1241F, 1956R, 3125R, 3223R, 769F, 1339F, 1468R, 1725F, 2649F, 2347F, 203R, 241R, 241F, 535F, 4856R, 666R. В результате была

расшифрована нуклеотидная последовательность гена, которая была названа *sit-tri*, имеющая ОРС длиной 4248 п.н. Предсказанная аминокислотная последовательность протяженностью 1415 а.о., кодируемая этой ОРС, имеет три предсказанных «зрелых» белка SIT: 1А – 435 а.о.; 1В – 445 а.о. и 2 – 440 а.о. (рис. 3Г). На следующем этапе было выявлено взаимное расположение генов *sit-td* и *sit-tri* относительно друг друга. Для этого был использован праймер 3106F, расположенный за 3'- концом гена *sit-td*, и праймер 241R из 5'- конца гена *sit-tri*. Также использовали второй способ для определения взаимного расположения мульти-*sit* генов *sit-td* и *sit-tri*, который описан ниже. Используя праймер 402F, расположенный в начале гена *sit-td*, и праймер 6663R, расположенный в участке между генами *sit-td* и *sit-tri*, была получена левая часть контига, содержащая ген *sit-td*. Правую часть контига с геном *sit-tri* получали, используя праймер 5056F из участка между генами *sit-td* и *sit-tri* и праймер (4856R), расположенный в конце гена *sit-tri* (рис. 4). В результате пошагового секвенирования ампликонов было показано, что гены мульти-*sit* *S. acus* subsp. *radians* содержат участок некодирующей ДНК (536 п.н.), расположенный перед геном *sit-td* (2520 п.н.) и содержащий прямой повтор длиной 483 п.н., который на 99,2 % идентичен копии, находящейся в межгенном участке (4180 п.н.) на расстоянии в 791 п.н. от гена *sit-td*. В результате общая протяженность полученного контига составила 11819 п.н. В межгенном участке на расстоянии 5885-6297 п.н. аннотирован фрагмент ретровирусного транспозона, гомологичного последовательности *gag-pol* *Thalassiosira pseudonana* (GenBank ACA60905), относящегося к семейству Ty1/Copia-like LTR (рис. 4).

С помощью RACE-ПЦР на суммарной РНК, выделенной из культуры *S. acus* subsp. *radians*, и подобранных концевых праймеров определены 5'- и 3'- концы транскриптов, соответствующих мульти-*sit* мРНК, и таким образом показано, что в транскрипте *S. acus* subsp. *radians* есть транскрипты обоих генов *sit-td* и *sit-tri*. Исходя из данных, полученных с помощью RACE-ПЦР, длина транскрипта гена *sit-td* составляет 2771 п.н. (позиции 381-3151), а транскрипта гена *sit-tri* – 4350 п.н. (7230-11579; позиции указаны по контигу GenBank KX345281.2).

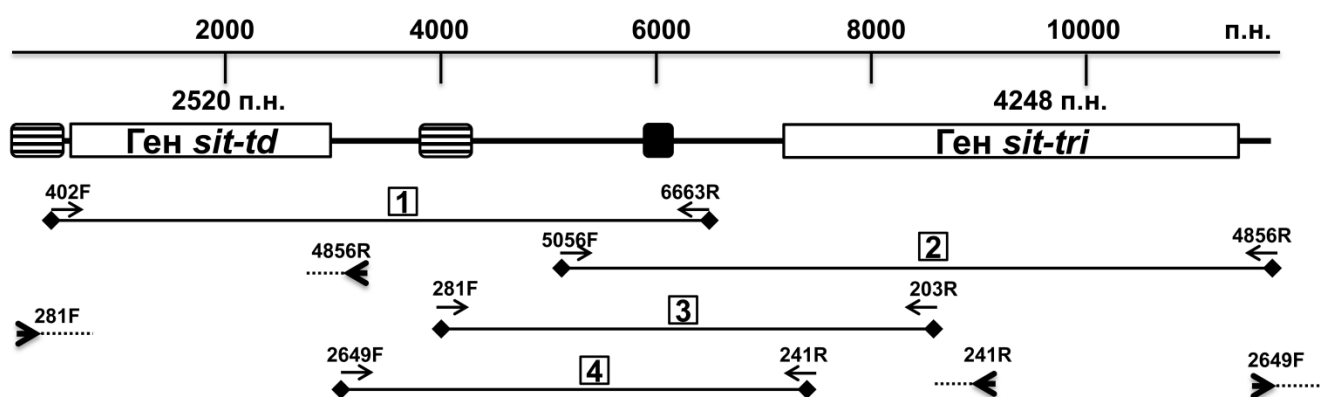


Рисунок 4. Схема перекрывания протяженных секвенированных фрагментов генов мульти-*sit* *S. acus* subsp. *radians*. Сплошные линии – секвенированные участки генов мульти-*sit*, полученные с геномной ДНК *S. acus* subsp. *radians*. → – места посадки специфичных праймеров. Прямоугольники с горизонтальной штриховкой – прямые повторы. Цифры в квадратах – ПЦР фрагменты, полученные с соответствующих пар праймеров. Черный квадрат – остаток ретровирусного транспозона (259 п.н.), гомологичный последовательности *gag-pol* *T. pseudonana* (GenBank ACA60905).

Обнаружение генов мульти-*sit* у диатомеи *S. ulna* subsp. *danica*

С целью обнаружения генов мульти-*sit* у *S. ulna* subsp. *danica* были выравнены нуклеотидные последовательности, кодирующие CMLD мотив и его окружение (CMLDFIN) у нескольких видов диатомей (*S. acus* subsp. *radians*, *Pseudo-nitzschia multiseries*, *Achnanthes exigua*, *Epithemia zebra*, *Phaeodactylum tricornutum*). Последовательность смысловой цепи ДНК, кодирующей CMLD и его окружение, с учетом вырожденности кодонов и консервативности позиций нуклеотидов была выбрана в качестве прямого праймера (UniCMLD_F). Аналогичным образом комплементарную антисмысловую последовательность данного участка выбрали в качестве обратного праймера (UniCMLD_R). В результате ПЦР с праймерами (UniCMLD_F- UniCMLD_R) и геномной ДНК *S. ulna* subsp. *danica* было получено несколько продуктов амплификации длиной около 1400 п.н., которые были зафиксированы клонированием, выделены индивидуальные продукты амплификации в составе плазмидной ДНК (рис. 5), секвенированы и затем сравнены с генами мульти-*sit* *S. acus* subsp. *radians*. Множественное выравнивание и филогенетический анализ позволили идентифицировать четыре группы клонированных ДНК (рис. 5).

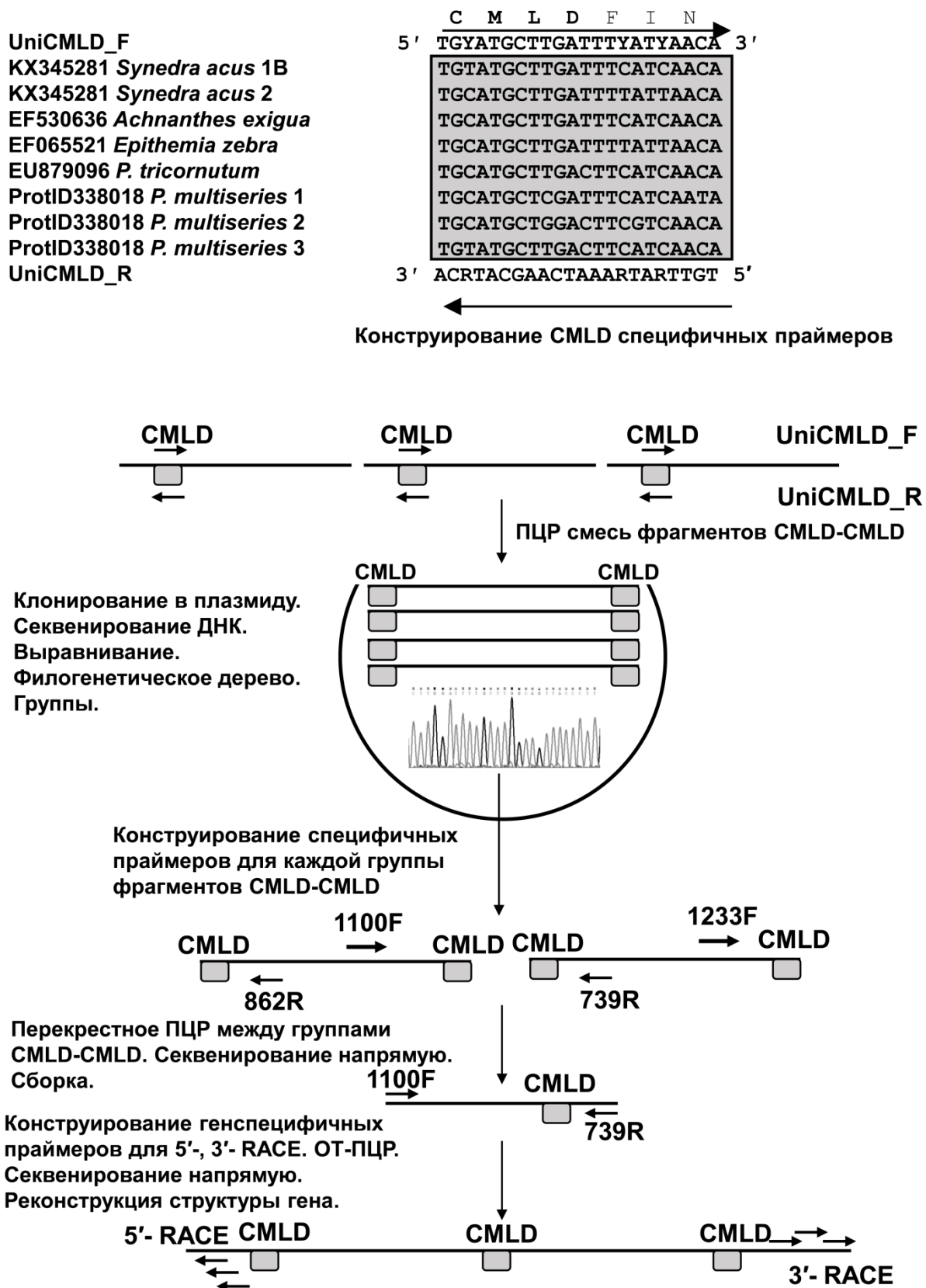


Рисунок 5. Блок-схема экспериментов по получению генов мульти-*sit* *S. ulna* subsp. *danica*. Горизонтальные стрелки – праймеры.

Для каждой из этих групп были сконструированы уникальные праймеры, направленные в сторону CMLD, для амплификации ДНК и установления связи между клонированными участками ДНК. Полученные ПЦР фрагменты секвенированы по методу Сэнгера. В результате было найдено два фрагмента нуклеотидной последовательности генов мульти-*sit*, каждый из которых кодировал полипептидную цепь с тремя предсказанными «зрелыми» белками.

С помощью RACE-ПЦР на суммарной РНК культуры *S. ulna* subsp. *danica* и выбранных концевых праймеров определены 5'- и 3'- концы транскриптов, соответствующие мульти-*sit* мРНК (**рис. 5**) и реконструированы последовательности генов мульти-*sit*.

Так как в геноме *S. acus* subsp. *radians* гены *sit-td* и *sit-tri* расположены на расстоянии 4180 п.н. друг от друга, то решено было проверить взаимное расположение мульти-*sit* генов в геноме *S. ulna* subsp. *danica*. Для этого использовали ПЦР с сочетанием концевых праймеров, направленных изнутри гена наружу – 1233F-862R и 3531F-172R. При этом ПЦР-продукт из внутренней части гена не может синтезироваться, а фрагменты ДНК, перекрывающие участок между генами мульти-*sit*, амплифицируются. Таким образом нами показано, что в геноме *S. ulna* subsp. *danica* имеется три мульти-*sit* гена, два из которых идентичны *sit1/sit2* (**рис. 6А**). Для того чтобы достоверно реконструировать полную последовательность участка, содержащего гены *sit1* и *sit2*, мы также амплифицировали полноразмерный ген *sit1* и межгенный участок (1657F-92R праймеры, **рис. 6Б**), перекрывающийся с ним межгенный участок и полноразмерный ген *sit2* (9288F-5561R праймеры, **рис. 6Б**), а также оба гена *sit1* и *sit2* (1657F-5561R праймеры, **рис. 6Б**). Показано, что гены *sit1* и *sit2* расположены на расстоянии друг от друга 5239 п.н. В результате общая протяженность полученного контига составила 14011 п.н. (**рис. 6Б**). Расположение гена *sit3* в геноме *S. ulna* subsp. *danica* относительно генов *sit1* и *sit2* остается неизвестным. С использованием RACE-ПЦР установлено, что протяженность транскриптов генов *sit1/sit2* составляет 4478 п.н. (позиции 1-4478; 9533-14010; позиции указаны по контигу, имеющему номер в базе данных GenBank MF971079), а *sit3* – 4361 п.н. (позиции 1-4361; позиции указаны по контигу, имеющему номер в базе данных GenBank MF971078).

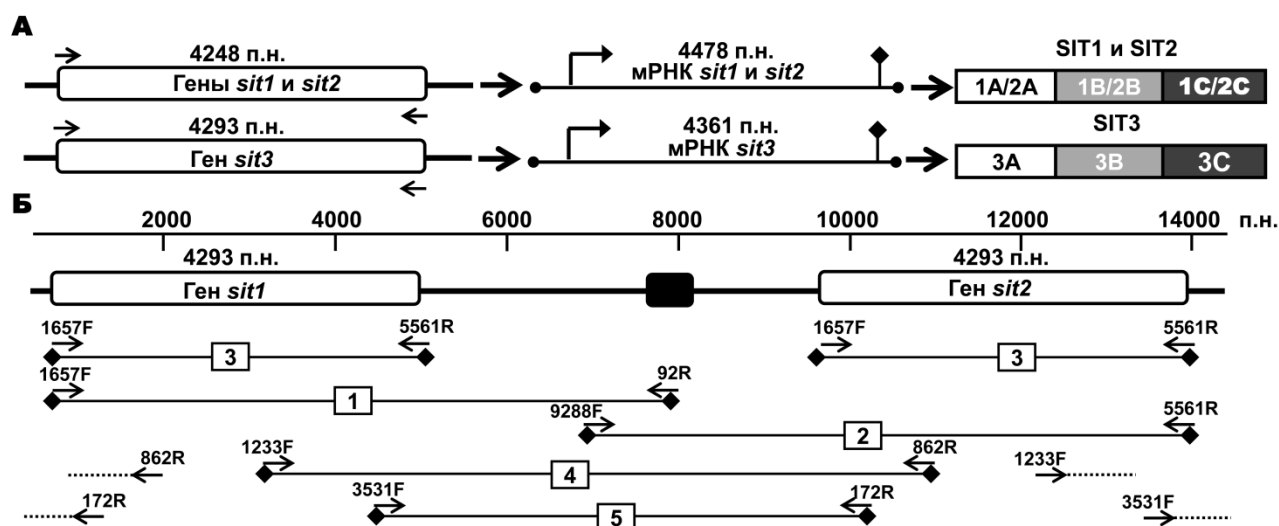


Рисунок 6. Расположение генов *sit* в геноме *S. ulna* subsp. *danica*. А – схемы генов *sit1*, *sit2* и *sit3*. Черные сплошные линии – секвенированные участки ДНК и РНК; прямоугольники – предсказанные аминокислотные последовательности SIT1, SIT2 и SIT3; блоки А, В, С – предсказанные «зрелые» белки SIT. Б – организация генов *sit1* и *sit2*. Прямоугольники – ОРС генов *sit1* и *sit2*. → места посадки праймеров; цифры в прямоугольниках – ПЦР фрагменты, полученные с пар праймеров: 1 – 1657F-92R (7259 п.н.); 2 – 9288F-5561R (7027 п.н.); 3 – 1657F-5561R (4417 п.н.); 4 – 1233F-862R (7818 п.н.); 5 – 3531F-172R (5464 п.н.). Черный прямоугольник – остаток ретровирусного транспозона (530 п.н.), гомологичный последовательности *gag-pol T. pseudonana* (GenBank ACA60905).

Так же как и для *S. acus* subsp. *radians* в межгенном участке на расстоянии 7539-8068 п.н. аннотирован фрагмент ретровирусного транспозона, имеющий гомологию с последовательностью *gag-pol T. pseudonana* (GenBank ACA60905), относящийся к семейству Ty1/Copia-like LTR (рис. 6Б). Это позволяет сделать предположение о дупликации генов мульти-*sit* у *S. acus* subsp. *radians* и *S. ulna* subsp. *danica* в результате ретротранспозиции.

Сравнительный анализ мульти-SIT белков

Сравнительным анализом 49 предсказанных аминокислотных последовательностей мульти-SIT белков из 17 видов диатомей и одного вида инфузории [Durkin *et al.*, 2012; Durkin *et al.*, 2016] показано, что на границах между «зрелыми» белками SIT расположены консервативные мотивы YQXDXVYL и DXDID (рис. 7), мотив DXDID обогащен аспарагиновой кислотой (D).

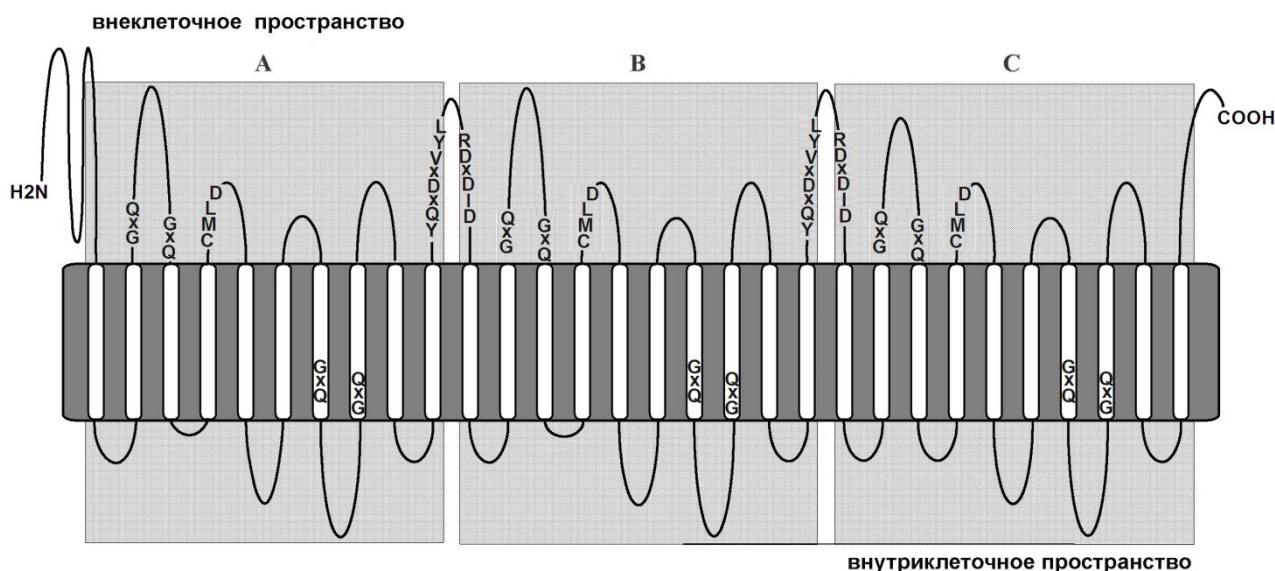


Рисунок 7. Схема расположения белков SIT1/2 *S. ulna* subsp. *danica* в цитоплазматической мембране, построенная с помощью программы TMpred. Трансмембранные домены (ТМ) – белые вертикальные линии; серые прямоугольники А, В, С – предсказанные «зрелые» белки SIT, полученные с помощью программы Pfam; GXQ, CMLD, YQXDXVYL, DXDID – консервативные мотивы. Ориентация внутрь/наружу клетки равновероятна.

В природе широко представлены белки, которые подвергаются посттрансляционному процессингу посредством разрезания специфическими протеазами; к примеру, каспазы, специфичные к мотивам с повторяющимися остатками аспарагиновой кислоты (D), регулируются у млекопитающих путём автопротеолитической активации [Cryns, Yuan, 1998].

Поскольку на конце последовательностей DXDID отсутствует пролин (P), который мог бы вносить излом во вторичную структуру белка и тем самым затруднять взаимодействие протеазы с данными аминокислотами, это может служить еще одним косвенным аргументом в пользу возможности протеолиза по определенным сайтам в процессе посттрансляционной модификации у диатомей. При разделении предсказанных аминокислотных последовательностей SIT-TD, SIT-TRI *S. acus* subsp. *radians* и SIT1/SIT2 и SIT3 *S. ulna* subsp. *danica* на «зрелые» белки SIT показано, что они имеют протяженность 392-446 а.о., а предсказанная молекулярная масса составляет 42,7-48,8 кДа. Длина предсказанных аминокислотных последовательностей белков SIT составляет 506-599 а.о. с предсказанной молекулярной массой 54,9-65 кДа. Из 49 разрезанных мульти-SIT белков получено 87 «зрелых»

белков SIT, границы которых определены с помощью онлайн-поиска доменов в базе данных Pfam.

Последовательность мотива CMLD, по мнению авторов, является высококонсервативной [Грачев и др., 2002] и может принимать участие в транспорте кремния путем взаимодействия кремниевой кислоты с мотивом CMLD и ионами Zn^{+2} [Grachev, Annenkov, Likhoshway, 2008]. Мотив CMLD присутствует у всех последовательностей кроме трех белков. Некоторые отдельные аминокислотные остатки характеризуются исключительной консервативностью. Например, остатки G в положениях 79, 192, 228, 363, 451, 497, остаток W – в положении 466, остатки L – в положениях 67, 119, 194, 459. Это указывает на то, что эти а.о. не изменились в процессе эволюции, и, вероятно, важны и участвуют либо в правильном фолдинге третичных структур, либо взаимодействуют с кремниевой кислотой.

ВЫВОДЫ

1. Определены нуклеотидные последовательности двух генов мульти-*sit* белков – транспортеров кремниевой кислоты у пресноводной пеннатной бесшовной диатомовой водоросли *S. acus* subsp. *radians*. Установлено, что каждый из этих генов мульти-*sit* продуцирует один транскрипт, предположительно кодирующий слитый белок, состоящий из двух или трех «зрелых» белков. Гены *sit-td* и *sit-tri* *S. acus* subsp. *radians* расположены в одной хромосоме на расстоянии 4180 п.н.

2. В геноме *S. ulna* subsp. *danica* определены нуклеотидные последовательности трех генов мульти-*sit*, два из которых *sit1* и *sit2* идентичны друг другу и расположены в одной хромосоме на расстоянии 5239 п.н. Предсказанные аминокислотные последовательности мульти-SIT белков *S. ulna* subsp. *danica* состоят из трех «зрелых» белков.

3. В участках между генами *sit-td* и *sit-tri* *S. acus* subsp. *radians* и генами *sit1* и *sit2* *S. ulna* subsp. *danica* аннотированы фрагменты ретровирусных транспозонов, имеющие гомологию с последовательностью *gag-pol* *Thalassiosira pseudonana*, наличие которых может свидетельствовать о дупликации генов *sit* в результате транспозиции.

4. Выявленный у 49 мульти-SIT белков из 17 видов диатомей консервативный мотив DXDID может являться мишенью протеолиза этих белков до отдельных «зрелых» белков SIT.

СПИСОК ПУБЛИКАЦИЙ ПО ТЕМЕ РАБОТЫ

Статьи в журналах, рекомендованных ВАК:

1. Галачьянц Ю.П., Захарова Ю.Р., Петрова Д.П., Морозов А.А., Сидоров И.А., **Марченков А.М.**, Логачева М.Д., Маркелов М.Л., Хабудаев К.В., Лихошвай Е.В., Грачев М.А. Определение нуклеотидной последовательности полного генома бесшовной пеннатной диатомеи *Synedra acus* subsp. *radians* из озера Байкал // Доклады Академии наук. 2015. Т. 461. С. 348–352.
2. **Марченков А.М.**, Бондарь А.А., Петрова Д.П., Хабудаев К.В., Галачьянц Ю.П., Захарова Ю.Р., Волокитина Н.А., Грачев М.А. Необычная конфигурация генов белка транспорта кремния у пресноводной пеннатной диатомеи *Synedra acus* subsp. *radians* // Доклады Академии наук. 2016. Т. 471. № 2. С. 238–240.
3. **Marchenkov A.M.**, Petrova D.P., Morozov A.A., Zakharova Y.R., Grachev M.A., Bondar A.A. A family of silicon transporter structural genes in a pennate diatom *Synedra ulna* subsp. *danica* (Kütz.) Skabitsch. // PLOS ONE. 2018. V. 13. No. 8. P. e0203161.

Прочие публикации:

4. **Марченков А.М.**, Бондарь А.А., Петрова Д.П., Хабудаев К.В., Грачев М.А. Изучение транскрипции генов транспорта кремниевой кислоты у пресноводной диатомовой водоросли *Synedra acus* // Тезисы VI Всероссийского с международным участием Конгресса молодых ученых-биологов «Симбиоз-Россия 2013» (Иркутск, 19-23 августа 2013 г.). Иркутск: Издательство Аспринт, 2013. С. 292–293.
5. Петрова Д.П., **Марченков А.М.** Белки-транспортеры кремния: долгий путь к открытию // Наука из первых рук. 2016. Т. 68. №. 2. С. 64–71.
6. **Марченков А.М.** Кластер генов белка транспорта кремниевой кислоты у пресноводной диатомовой водоросли *Synedra acus* subsp. *radians* // Тезисы VII Международной школы молодых ученых по молекулярной генетике «Геномика и биология живых систем» (Звенигород, 14-18 ноября 2016 г.). Москва: ИМГ РАН, 2016. С. 7.
7. **Марченков А.М.**, Петрова Д.П., Морозов А.А., Бондарь А.А., Грачев М.А. Мультиплицированные гены транспорта кремниевой кислоты (*sit*) у диатомовой водоросли *Synedra ulna* subsp. *danica* //

Тезисы IV международной конференции молодых ученых: биотехнологов, молекулярных биологов и вирусологов (Научноград Кольцово, 24-26 октября 2017 г.). Новосибирск: НГУ, 2017. С. 262–265.

Благодарности. Автор глубоко признателен своему научному руководителю академику М.А. Грачеву за постановку цели и задач, к.х.н. А.А. Бондарю (ИХБФМ СО РАН) и к.б.н. Д.П. Петровой (ЛИН СО РАН) за неоценимую помощь во время работы, интерпретации и обсуждении результатов, к.б.н. А.А. Морозову (ЛИН СО РАН) за помощь в анализе полученных данных, ценные советы и замечания, к.б.н. Ю.Р. Захаровой и Н.А. Волокитиной (ЛИН СО РАН) за предоставления биомассы клеток, к.б.н. Ю.П. Галачьянцу (ЛИН СО РАН) за предоставление полногеномных данных *S. acus* subsp. *radians*, К.В. Хабудаеву (ЛИН СО РАН) за аннотацию генов, д.б.н. Е.В. Лихошвай за ценные замечания и помощь в оформлении диссертации. Автор выражает искреннюю благодарность сотрудникам отдела ультраструктуры клетки (ЛИН СО РАН) и ЦКП «Геномика» (ИХБФМ СО РАН).

Работа выполнена в рамках тем бюджетного финансирования № 0345-2014-0006 «Экспериментальные исследования геномов и протеомов», 0345-2016-0005 «Экспериментальные исследования геномов и протеомов биоты пресноводных экосистем» и поддержана программой Президиума РАН «Молекулярная и клеточная биология», проект № 0345-2015-0031.

МАРЧЕНКОВ Артём Михайлович
СИСТЕМА ГЕНОВ ТРАНСПОРТА КРЕМНИЕВОЙ КИСЛОТЫ
У ПРЕДСТАВИТЕЛЕЙ ДИАТОМОВЫХ ВОДОРΟΣЛЕЙ РОДА
SYNEDRA EHRENBURG

Автореф. дисс. на соискание учёной степени
кандидата биологических наук

Подписано в печать _____.

Формат 60x84/16. Усл. печ. л. 1. Тираж 100 экз. Заказ № ____.

Издательство Института географии им. В.Б. Сочавы СО РАН

664033 г. Иркутск, ул. Улан-Баторская, 1